

# **ECONOMIC THEORY, APPLICATIONS AND ISSUES**

**Working Paper No. 52**

**Quantitative Impacts of Teaching  
Attributes on University TEVAL Scores  
and their Implications**

**By**

**Mohammad Alauddin**

**and**

**Clem Tisdell**

**April 2009**



ISSN 1444-8890

**ECONOMIC THEORY, APPLICATIONS AND ISSUES  
(Working Paper)**

**Working Paper No. 52**

**Quantitative Impacts of Teaching Attributes  
on University TEVAL Scores and  
their Implications\***

**by**

**Mohammad Alauddin<sup>†</sup>**

**And**

**Clem Tisdell<sup>‡</sup>**

**April 2009**

© All rights reserved

---

\* This a revised and improved version of Working Paper No. 42 in this series.

<sup>†</sup> School of Economics, The University of Queensland, Brisbane, Australia 4072, Email:  
[m.alauddin@economics.uq.edu.au](mailto:m.alauddin@economics.uq.edu.au)

<sup>‡</sup> School of Economics, The University of Queensland, St. Lucia Campus, Brisbane QLD 4072, Australia  
Email: [c.tisdell@economics.uq.edu.au](mailto:c.tisdell@economics.uq.edu.au)

WORKING PAPERS IN THE SERIES, *Economic Theory, Applications and Issues*, are published by the School of Economics, University of Queensland, 4072, Australia.

For more information write to Professor Clem Tisdell, School of Economics, University of Queensland, St. Lucia Campus, Brisbane 4072, Australia or email [c.tisdell@economics.uq.edu.au](mailto:c.tisdell@economics.uq.edu.au)

# **Quantitative Impacts of Teaching Attributes on University TEVAL Scores and their Implications**

## **ABSTRACT**

This article uses a large sample of completed student evaluation forms of university teaching to quantify the impacts of student evaluation of teaching (SET) attributes on teaching effectiveness (TEVAL) scores. Despite much criticism of and support for TEVAL scores measuring teaching effectiveness, detailed quantitative studies of the relationship between instructional attributes and TEVAL scores are lacking.

This study helps to fill this gap. Results suggest that the relative influence of teaching attributes on TEVAL scores varies with the level of the course. While students' perceptions of how well the coursework is organized, explained and presented have large positive impacts on TEVAL scores at all levels, their relative importance varies with the level of the class. Furthermore, the SET attribute "emphasis on thinking rather than memorizing" has little or no substantive impact on TEVAL scores. An implication is that a lecturer stressing this aspect does little to increase her/his TEVAL score. Furthermore, lecturers wishing to raise their TEVAL scores should vary their relative emphasis on different teaching attributes according to the class level. A feature of this study is its use of individual student responses rather than class averages. Therefore, it accounts for all the information provided by the data.

# **Quantitative Impacts of Teaching Attributes on University TEVAL Scores and their Implications**

## **1. INTRODUCTION**

The practice of student evaluation of teaching (SET) has shown extraordinary growth since the early 1970s (Seldin 1998). Student evaluations are currently considered the most important, and sometimes are the sole, measure of an instructor's teaching ability (Wilson 1998, p. A12). At the university level, SET is invariably used in promotion or tenure decisions as the most important indicator of an instructor's pedagogical performance or teaching effectiveness. While there has been much criticism of and support for TEVAL scores as a measure of teaching effectiveness (see for example, Felton et al. 2008), detailed quantitative studies of the relationship between instructional attributes used in SET and TEVAL scores are lacking. Therefore, this study helps to fill a gap in the literature. A significant feature of this study is its use of individual student responses and represents a departure from the aggregative type of analysis relying on class averages. For one thing, a disaggregated analysis involving individual data can capture the underlying heterogeneity within a group of respondents while analysis based on class averages mask it.

More specifically, results reported and analyzed in this paper, are based on data obtained from 2467 SET forms completed for various economics courses at The University of Queensland, a large Australian university. The practice followed at The University of Queensland regarding TEVAL scores is typical of that followed by most universities in higher income countries. The SET questionnaires use instructional attributes for evaluation purposes. This paper uses these data to:

1. identify instructional attributes that result statistically significant and large variations in TEVAL scores and those that do not;

2. investigate whether their effects on TEVAL vary across different levels of courses; and
3. examine the implications of major findings stemming from the results.

Section 2 outlines the main features of the data. Section 3 presents and discusses the empirical results. Section 4 provides and examines results of responsiveness of TEVAL to changes in instruction attributes. Section 5 concludes the paper.

## **2. THE DATA: AN INTERPRETIVE ANALYSIS**

The basic data for this study are from the SET surveys of nine economics courses that include four large second and two large third level undergraduate courses and three large postgraduate courses at The University of Queensland. They are for the period 2000 to 2006 and are based on 2467 completed SET forms of which 1573 refer to the undergraduate samples across six courses at two levels while 894 relate to three introductory postgraduate courses.

The data do not meet the criterion of strict randomness in the sense that they could not be selected at random. This is because many university staff members are sensitive to letting others use their *TEVAL* records for research. Nevertheless, the data used in this study relate to a large range of courses – including large-sized second and third level undergraduate and postgraduate courses. These courses displayed considerable diversity in their student populations typified by variations academic background, degree destination, and English language competency of students. The data employed in this study that this study relates to different cohorts of students who rated different aspects (components) of an instructor's teaching. The cohorts are from independent populations and the data are not longitudinal.

Note that The University of Queensland requires all instructors to collect *TEVAL* data. However, the collected data represent the responses from only those students who are present in the class on the day *TEVAL* surveys take place. Thus, not every student has an equal chance of appearing in the data. Those students who are less likely to attend classes

are under-represented. they may have chosen not to attend as frequently as others, they probably do so for a variety of reasons including ‘lack of interest in lectures’, work and family commitments, alternative forms of access to learning resources e.g. *eLearning*, electronic communications and so on. For one thing, those students attend and determine an instructor’s *TEVAL* score.

Table 1 provides the codes and definitions of dependent and independent variables.

**Table 1:** Definitions and description of teaching attributes used in SET data

<b>Variable Code</b>	<b>Description</b>
<i>TEVAL</i>	<b>Dependent variable:</b> All things considered, how would you rate this lecturer’s overall effectiveness as a university teacher? (1 – very poor, 5 – outstanding)
<b>Independent variables:</b> Instructor attributes ((1 – strongly disagree, 5 – strongly agree)	
<i>ORGANIZE</i>	The lecturer produced classes that were well organized
<i>PRESENT</i>	The lecturer presented material in an interesting way
<i>FEEDBACK</i>	The lecturer gave adequate feedback on my work
<i>RESPECT</i>	The lecturer treated students with respect
<i>KNOWWELL</i>	The lecturer seemed to know the subject well
<i>ENTHUSM</i>	The lecturer communicated her/his enthusiasm for the subject
<i>THINKMEM</i>	The lecturer emphasized thinking rather than memorizing
<i>EXPLAIN</i>	The lecturer gave explanations that were clear
<i>CONSULT</i>	The lecturer was available for consultation
<i>LSKILLS</i>	The lecturer helped to improve my learning skills

Given the ordinal nature of the data, median and mode, not mean, are the appropriate measures of central tendency. The educational literature and the administrators alike routinely use the mean rather than median or mode even though it is not correct to do so from a statistical point of view in case of ordinal data (Selvanathan et al. 2006, p. 125). This is because even though one could code an “outstanding” TEVAL score as “5”, it does not necessarily mean that it is five times as good as a “very poor” TEVAL score, coded as “1”.

The descriptive statistics (not reported here for brevity) reveal that the distributions of *TEVAL* scores and other instructional attributes are considerably 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 is in the third quartile ( $Q_3$ ) while the first quartile ( $Q_1$ ) without exception was located the 3-4 range.

We applied a statistical test to determine whether the distributions for TEVAL and other attributes differed between postgraduate and undergraduate samples and between the two levels of the undergraduate program. A two-sample Kolmogorov-Smirnov test indicates that *TEVAL* distributions for the two undergraduate samples are significantly different ( $p$ -value =0.000). The same test also reveals that the distributions for level 3 undergraduate (UG3) and the postgraduate (PG) samples do not differ significantly ( $p$ -value =0.992). However, as expected the distributions relating the lower undergraduate sample and the postgraduate one are significantly different ( $p$ -value =0.000). Note that a transition from the lower to the upper level undergraduate courses leads to decline in the standard deviation of the distribution of *TEVAL* scores as a student moves upwards. All else being equal, 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.

Several factors may explain this 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 comprehension of the subject. This may be reflected in their *TEVAL* scores. Furthermore,



sorting is likely to occur. Those students who are less enthusiastic or less able to cope with a subject are less likely to continue with it in later years than those who are more capable and enthusiastic. This, in all probability, will be reflected in the distribution of the *TEVAL* scores. However, further research is needed to identify more accurately 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 lower level classes should be adjusted to be comparable with scores of those teaching higher-level courses.

Pearson correlation coefficients between *TEVAL* and the remaining variables indicate significant positive correlations. However, it can also be seen that for the entire sample data *EXPLAIN*, *PRESENT*, *ORGANIZE* and *LSKILLS* 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. The results are not presented here but are available upon request.

### **3 EMPIRICAL RESULTS**

A large body of literature recognizes that linear regression is inappropriate when the dependent variable is categorical, especially if it is qualitative. Consider a customer survey where responses are coded 1 (worst/strongly disagree), 2, 3, 4 or 5 (best/strongly agree). Green (2000, p.875) states, “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”. The appropriate theoretical model in such a situation is the ordered probit model (see for example, Greene 2000). Since McKelvey and Zovoina (1975), these models have been widely used as a methodological framework for analyzing ordered data.

The dependent variable, *TEVAL*, is coded from zero to four. Table 2 presents two estimated equations for the two levels of undergraduate courses and one equation for the postgraduate sample using all the instruction attributes (as perceived by the students) listed in Table 1 as independent variables. All the equations include course dummies. The dummy variable is set a value of zero for the course with the lowest mean *TEVAL* score.

Thus, the course dummies for UG2 C1, UG3 C1 and PG C1 assume zero values. All other course dummies assume a value of unity.

The values of the pseudo- $R^2$  ranges between 0.38 and 0.51 indicating reasonable fits for all the models. Since the traditional  $R^2$  is a 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 Moffat, 2002; Greene 2000, p. 683).

An inspection of the results suggests that:

- For undergraduate level 2 (UG2), the course dummy variables are not statistically significant implying course-neutrality. All but two of the ten independent variables are statistically significant. Furthermore, their magnitudes show that they have substantial impact on *TEVAL*. The coefficients of *RESPECT* and *ENTHUSM* are not statistically significant. Nor are they numerically substantive.
- For UG3, seven out of the ten independent variables appear to be statistically significant (the exceptions being *RESPECT*, *ENTHUSM* and *THINKMEM*). The statistical significance of the coefficient of the dummy variable indicates that determinants of *TEVAL* vary significantly across courses.
- For postgraduate (PG), estimated coefficients of all but three independent variables (*RESPECT*, *KNOWWELL* and *CONSULT*) appear significant. The negative sign of the coefficient of *RESPECT* appears to be counter-intuitive. Of the two course dummies, the one for PG C2 is statistically significant.

Two common concerns often expressed are that: (1) serious multicollinearity problem renders the empirical results less useful; (2) the results are not robust (consistent). The present study tested for both of these problems.

The variance inflation factors (VIF) set out in Table 3 suggest that multicollinearity is not an issue in the present study as none of the VIFs exceed even the conservative threshold of 5 (Snee 1973). In fact all the VIFs are below 3.

To examine the robustness of the estimates with different sample sizes, standard errors of the parameters were estimated using the bootstrap approach (Efron & Tibshirani, 1993). The results suggest that the present samples may reflect the population as the bootstrapped estimates, using the common choices of 1000 and 2000 repetitions are very similar to those obtained from the original regressions (see Table 2). Thus, the results of the ordered probit analysis are robust.

Against the above background, one can analyze the implications of statistically significant dummy variables by comparing the probabilities that result when this variable takes its two different values with those that occur with all other variables held at their mean values (Greene 2000, p. 879). In light of this, the instructor in UG3 C2 (relative to UG2 C2) has a 5.89 per cent lower and an 8.81 per cent higher probability of getting a *TEVAL* score of 4 and 5 respectively. Likewise, the instructor in PG2 C2 (compared to UG2 C2) has an 8.49 per cent lower and a 12.73 per cent higher chance of *TEVAL* scores of 4 and 5 respectively. Therefore, the higher the level of the class the more likely is a lecturer to obtain a larger *TEVAL* score for the same rating of attributes.

**Table 2:** Results of ordered probit analysis and bootstrap estimates of overall perceived teaching effectiveness score (*TEVAL*) by perceived teaching for different levels and programs

Variable	Coefficient								
	UG2 (Original)	UG2 (BT1000)	UG2 (BT2000)	UG3 (Original)	UG3 (BT1000)	UG3 (BT2000)	PG (Original)	PG (BT1000)	PG (BT2000)
CONSTANT	4.631 <sup>a</sup>	4.554 <sup>a</sup>	4.554 <sup>a</sup>	5.167 <sup>a</sup>	4.869 <sup>a</sup>	4.869 <sup>a</sup>	3.030 <sup>a</sup>	3.030 <sup>a</sup>	3.030 <sup>a</sup>
<i>ORGANIZE</i>	0.389 <sup>a</sup>	0.389 <sup>a</sup>	0.389 <sup>a</sup>	0.625 <sup>a</sup>	0.625 <sup>a</sup>	0.625 <sup>a</sup>	0.454 <sup>a</sup>	0.454 <sup>a</sup>	0.454 <sup>a</sup>
<i>PRESENT</i>	0.425 <sup>a</sup>	0.425 <sup>a</sup>	0.425 <sup>a</sup>	0.384 <sup>a</sup>	0.383 <sup>a</sup>	0.383 <sup>a</sup>	0.223 <sup>a</sup>	0.223 <sup>a</sup>	0.223 <sup>a</sup>
<i>FEEDBACK</i>	0.045	0.045	0.045	0.256 <sup>a</sup>	0.255 <sup>a</sup>	0.255 <sup>a</sup>	0.206 <sup>a</sup>	0.208 <sup>a</sup>	0.208 <sup>a</sup>
<i>RESPECT</i>	0.128 <sup>b</sup>	0.128 <sup>b</sup>	0.128	0.158	0.157	0.157	-0.041	-0.042	-0.042
<i>KNOWWELL</i>	0.257 <sup>a</sup>	0.257 <sup>a</sup>	0.257 <sup>a</sup>	0.192	0.192	0.192	0.087	0.087	0.087
<i>ENTHUSM</i>	0.065	0.065	0.065	0.130	0.130	0.130	0.181 <sup>b</sup>	0.181	0.181
<i>THINKMEM</i>	0.119 <sup>b</sup>	0.119	0.119	0.081	0.081	0.081	0.142 <sup>b</sup>	0.142 <sup>b</sup>	0.142
<i>EXPLAIN</i>	0.464 <sup>a</sup>	0.464 <sup>a</sup>	0.464 <sup>a</sup>	0.442 <sup>a</sup>	0.442 <sup>a</sup>	0.442 <sup>a</sup>	0.208 <sup>a</sup>	0.206	0.206
<i>CONSULT</i>	0.153 <sup>a</sup>	0.153 <sup>b</sup>	0.153 <sup>b</sup>	0.267 <sup>a</sup>	0.268 <sup>a</sup>	0.268 <sup>a</sup>	0.017	0.017	0.017
<i>LSKILLS</i>	0.376 <sup>a</sup>	0.376 <sup>a</sup>	0.376 <sup>a</sup>	0.264 <sup>a</sup>	0.264 <sup>a</sup>	0.264 <sup>a</sup>	0.403 <sup>a</sup>	0.403 <sup>a</sup>	0.403 <sup>a</sup>
<i>COURSE 2</i>	0.048	0.048	0.048	-0.298 <sup>b</sup>	-0.298	-0.298	0.339 <sup>a</sup>	0.339 <sup>a</sup>	0.339 <sup>a</sup>
<i>COURSE3</i>	-0.076	-0.076	-0.076	-	-	-	-0.024	-0.024	-0.024
<i>COURSE 4</i>	-0.102	-0.102	-0.102	-	-	-	-	-	-
$\mu_1$	1.455 <sup>a</sup>	1.455 <sup>a</sup>	1.455 <sup>a</sup>	2.262 <sup>a</sup>	2.262 <sup>a</sup>	2.262 <sup>a</sup>	1.447 <sup>a</sup>	1.447 <sup>a</sup>	1.447 <sup>a</sup>
$\mu_2$	3.778 <sup>a</sup>	3.779 <sup>a</sup>	3.779 <sup>a</sup>	4.684 <sup>a</sup>	4.684 <sup>a</sup>	4.684 <sup>a</sup>	3.376 <sup>a</sup>	3.376 <sup>a</sup>	3.376 <sup>a</sup>
$\mu_3$	6.045 <sup>a</sup>	6.045 <sup>a</sup>	6.045 <sup>a</sup>	7.139 <sup>a</sup>	7.139 <sup>a</sup>	7.139 <sup>a</sup>	5.237 <sup>a</sup>	5.237 <sup>a</sup>	5.237 <sup>a</sup>
$\chi^2(10)$	1219.51 <sup>a</sup>	1219.51 <sup>a</sup>	1219.51 <sup>a</sup>	592.18 <sup>a</sup>	592.18 <sup>a</sup>	592.18 <sup>a</sup>	732.24 <sup>a</sup>	732.24 <sup>a</sup>	732.24 <sup>a</sup>
<i>N</i>	929	929	929	490	490	490	823	823	823
<i>Pseudo R</i> <sup>2</sup>	0.48	0.48	0.48	0.51	0.51	0.51	0.38	0.38	0.38

**Notes:** a, b and c respectively represent original and bootstrap estimates with 1000 (BT1000) and 2000 (BT2000) repetitions. a and b respectively represent 1 and 5 per cent levels of significance (two-tail). UG2, UG3, and PG respectively refer to undergraduate level 2, undergraduate level 3 and postgraduate samples.

**Table 3:** Results of variance inflation factor (VIF) tests of multicollinearity.

Variable	Variance inflation factor (VIF)		
	UG2	UG3	PG3
<i>ORGANIZE</i>	2.331	2.320	2.268
<i>PRESENT</i>	2.513	2.632	2.342
<i>FEEDBACK</i>	1.779	1.866	2.101
<i>RESPECT</i>	1.949	2.045	2.174
<i>KNOWWELL</i>	2.695	2.564	2.915
<i>ENTHUSM</i>	2.132	2.105	2.976
<i>THINKMEM</i>	2.198	2.398	2.732
<i>EXPLAIN</i>	2.674	2.725	2.342
<i>CONSULT</i>	1.471	1.645	2.028
<i>LSKILLS</i>	2.392	2.725	2.841

*Note:* All the VIFs are far below 5, which is the critical level for presence of multicollinearity (see for example, Snee1973).

Apart from the effects of dummy variables discussed above, based on the originally estimated equations presented in Table 2, the four most important factors that affect *TEVAL* can be identified in order of the magnitudes of their coefficients and are set out in Table 4. These four factors are *EXPLAIN*, *ORGANIZE*, *PRESENT* and *LSKILLS*. However, based on numerical magnitudes, their rankings vary across samples. For example:

- *EXPLAIN* and *PRESENT* are the two most important factors for the UG2 sample followed closely by *ORGANIZE* and *LSKILLS*.
- In the UG3 sample, *ORGANIZE* is by far the most important factor, while the next two factors *EXPLAIN* and *PRESENT* are close to each other. *CONSULT* ranks a distant fourth.
- For the PG sample, *ORGANIZE* is the most important instructional attribute with *LSKILLS* not far behind. Both *PRESENT* and *EXPLAIN* exert relatively smaller influence on *TEVAL* (about half or less than half of *ORGANIZE* and *LSKILLS*).

**Table 4:** Four most important factors influencing *TEVAL* in order of the magnitudes of their coefficients by level and program

<b>Ranking</b>	<b>UG2</b>	<b>UG3</b>	<b>PG</b>
1 (Highest)	<i>EXPLAIN</i> (0.464)	<i>ORGANIZE</i> (0.625)	<i>ORGANIZE</i> (0.454)
2	<i>PRESENT</i> (0.425)	<i>EXPLAIN</i> (0.442)	<i>LSKILLS</i> (0.403)
3	<i>ORGANIZE</i> (0.389)	<i>PRESENT</i> (0.384)	<i>PRESENT</i> (0.223)
4 (Lowest)	<i>LSKILLS</i> (0.376)	<i>CONSULT</i> (0.267)	<i>EXPLAIN</i> (0.208)
<b>Range</b>	<b>0.088</b>	<b>0.358</b>	<b>0.246</b>

#### 4 SOME FURTHER ANALYSIS OF THE EMPIRICAL RESULTS

In light of the preceding discussion, this section considers change in the probability of a *TEVAL* score when the perceived attributes used for the estimated equations in Table 2 show an increase. We start with a base case where all attributes are given a rating of 4 and estimate the corresponding probability of an instructor getting a student rating of 5. It can be seen from Table 6 that in the base case, the estimated probability of the instructor being rated 5 is appreciably higher in case of the PG sample than in either of the two UG samples. Thus, an instructor with all perceived attributes set at 4, has about a 15 and a 13 per cent chances of a rating of 5 for the UG sample while the base case yields over a 33 per cent chance of a student rating of 5 for the PG sample.

Let us now consider the probability of an instructor getting a 5 when ratings of all attributes are increased from 4 to 5 (Table 6). The probability of getting a 5 for teaching effectiveness is most influenced by *ORGANIZE*, *EXPLAIN*, *PRESENT*, and *LSKILLS*. Predicted probabilities for the remaining six perceived attributes were estimated but are not reported in Table 5 for the sake of brevity.

The degree of variation in the *TEVAL* score differs across levels and programs. For example:

- In case of UG2, increasing the score of *ORGANIZE* from 4 to 5, *ceteris paribus* increases the probability of *TEVAL* = 5 from 14.9 to 25.7 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 28.2, 26.9 and 25.3 per cent in a *TEVAL* score from 4 to 5 from the base level of 14.9 per cent.

- For UG3, increasing *ORGANIZE* from 4 to 5 the probability of *TEVAL* is likely to increase from 13.2 to 31.1 per cent while the same margin of change in *PRESENT*, *EXPLAIN*, and *LSKILLS* is likely to increase the probability of *TEVAL* respectively to 23.1, 24.9 and 19.7 per cent.
- For the PG sample a transition from 4 to 5 in respect of *ORGANIZE* is likely to increase the probability of a *TEVAL* score of 5 from a base level 33.3 per cent to 50.9 per cent. A transition from 4 to 5 in respect of *PRESENT*, *EXPLAIN*, and *LSKILLS* is likely to increase the probability of *TEVAL* score of 5 respectively to 41.9, 41.1 and 48.9 per cent from the same base level of 33.3 per cent.

**Table 5:** Variations in predicted probability (measured in percentage) of *TEVAL* due to increases in the rating of four most influential instructional attributes from 4 to 5 *ceteris paribus*

Probability	Base case*	<i>ORGANIZE</i>	<i>PRESENT</i>	<i>EXPLAIN</i>	<i>LSKILLS</i>
TEVAL=4 (UG2)	74.1	69.0	68.2	67.3	69.3
TEVAL=5 (UG2)	14.9	25.7	26.9	28.2	25.3
TEVAL=4 (UG3)	77.8	66.4	72.6	71.3	74.9
TEVAL=5 (UG3)	13.2	31.1	23.1	24.9	19.7
TEVAL=4 (PG)	59.0	46.1	58.8	59.9	55.8
TEVAL=5 (PG)	33.3	50.9	41.9	41.1	48.9

\*All attributes =4.

A particular feature of this paper is that it provides useful information that can help improve perceived *TEVAL* score. For example, an instructor when teaching:

1. lower undergraduate students, needs to focus more or less equally on the four key attributes: *EXPLAIN*, *PRESENT*, *ORGANIZE* and *LSKILLS*;
2. upper undergraduate students should concentrate most on *ORGANIZE* followed by *EXPLAIN* and *PRESENT*; and
3. postgraduate students should focus most on *ORGANIZE* followed closely by *LSKILLS*.

One disturbing and surprising, feature of these results is that a key factor, *THINKMEM*, is statistically significant in only two out of three samples (UG2 and PG). More critically, in terms of numerical significance, it ranks 7<sup>th</sup> and 8<sup>th</sup> respectively for the PG and UG2 samples in the list of ten core instructional attributes.

In a limited number of cases (n = 677; 398 for 2005 and 279 for 2006; 224 for UG2; 128 for UG3 and 325 for PG) information was available on an additional item: “I have achieved the graduate attributes which the course aimed to develop (e.g. oral/written communication, team work, critical thinking, problem solving) (GRADATTR)”. The ordered probit analysis did not find any statistically significant coefficients across any of the samples.

The University of Queensland places significant emphasis on their students acquiring key graduate attributes given that the electronic course profiles of necessity have to display which learning activities map into which graduate attributes. While they may vary across courses and disciplines, developing critical judgment and analytical abilities form key elements of these attributes. *THINKMEM* is the only one core element in the list of ten core instructional attributes (Table 1). One could conceivably consider *THINKMEM* and/or *GRADATTR* as the variables representing The University of Queensland’s key graduate goal. However, given the lack of statistical or numerical significance of the coefficients in any of the samples, there appears to be a dichotomy. It is highly probable that improving *TEVAL* based on *SET* may result in reduced efforts being given to *THINKMEM* or to *GRADATTR*. On the face of the evidence presented here, it would be natural for a lecturer to de-emphasize these attributes and to concentrate on strengthening the other attributes (mentioned above) which have a larger impact on *TEVAL* scores.

## **5 CONCLUSIONS**

Whether or not lecturers provided well-organized lectures was numerically important in its impact on *TEVAL* scores at all class levels as well as statistically significant. Whether or not lecturers were perceived to explain their material well and present it



well was also important in its impact on TEVAL scores but less so for higher-level classes than lower level ones. THINKMEM only has a minor impact on TEVAL scores at all class levels. For some classes where GRADATTR was available in the questionnaire, it also turned out to have little influence on TEVAL scores. Furthermore, it was found that for the same ratings of teaching attributes TEVAL scores are lower for lower level classes. This is one indication that it is difficult to get high TEVAL scores in lower level classes than in higher-level classes.

An instructor looking at her/his SET results, in conjunction with the empirical findings of this paper, should have a strong message about how to improve her/his *TEVAL* score, and a department chair would gain a strong impression of the teaching strengths and weaknesses of an instructor. An instructor with a low *TEVAL* could demonstrate strength in *THINKMEM* and improving learning skills of students both of which are extremely important pedagogical responsibilities of an instructor. It appears to be the case for UG2 C1. Our results are consistent with the finding that “the literature consistently shows that high level cognitive skills material plays little role in raising SET scores, and by reducing clarity may actually lower such scores. ...” (Everett 1977, pp. 101-2).

The lack of statistical significance or numerical impact of THINKMEM is disturbing because high *TEVALs* can be achieved at the cost of some critically important factors in teaching and learning. However, one needs to be reminded though that the ordered probit analysis itself suffers from limitations. One such limitation, for example, is that the relationship it detects in relation to each of its component variables is only monotonic. Furthermore, it involves an additive function but a multiplicative relationship can sometimes be important. The nature of the mathematical relationship affects the possible results (associations) obtained. For example, *THINKMEM* could improve *TEVAL* scores up to a point but cause these to decline if the instructor makes the students think too much. However, the general relationship may be positive. Again, the scaling of instructor attributes such as *THINKMEM* can vary across students because the quantitative measurement of these is left open.

One key pragmatic question that needs to be asked is, how can instructors raise students' stated scores for key variables (such as *ORGANIZE* and *PRESENT*) that impact significantly and substantially on their *TEVAL* scores. One way might be to make 'objective' improvements in key pedagogic variables that this paper has identified on the expectation that they will be reflected in higher stated scores by students. A problem, however, is that these relationships appear to be not well documented quantitatively. A second approach would be to keep the nature of the instruction constant but spend more time trying to convince students how well the instruction has been organized, how well the teaching materials have been presented and so on. This involves promotional effort by the instructor in relation to the key variables identified in this paper. Unfortunately, neither approach may result in 'good' teaching practice.

In this context, it is worth mentioning a view expressed in a recent study by Bursdal and Harrison (2008), believe in the validity of SETE (student evaluation of teaching effectiveness). They point out (p.574) that

“SETE instruments measure student's *attitudes* to teaching effectiveness, not necessarily teaching effectiveness *per se*. ... A reliable and valid measure of a particular group's perceptions regarding effectiveness is not the same thing as having a valid measure of effective teaching” and recommend the development of teaching development of teaching portfolios for university academics. They further state, “the portfolio should have many indicators of their teaching performance. Student evaluations should be just one of them”.

Our results support this point of view.

## **ACKNOWLEDGEMENTS**

The authors wish to thank Chris O'Donnell and Nghiem Hong Son for assistance with the econometric analysis, Temesgen Kifle and Asad Khan for extensive consultation with the statistical aspects of the paper and Farhanaz Alauddin for data entry. We

gratefully acknowledge our (anonymous) colleagues for access to their *TEVAL* data for this research. Usual *caveats* apply.

## REFERENCES

- Bursdal, C.A. and Harrison, P.D. 2008. Further evidence supporting the validity of both a multidimensional profile and an overall evaluation of teaching effectiveness. *Assessment and Evaluation in Higher Education*. 33: 567-76.
- Crouchley, R. and R.B. Davies. 1999. A comparison of population average and random effect models for the analysis of longitudinal count data with base-line information. *Journal of Royal Statistical Society A*. 162: 331-47.
- Daykin, A. R. and Moffatt, P. G. 2002. Analyzing ordered responses: A review of the ordered probit model. *Understanding Statistics*, 1: 157-66.
- Efron, B. and Tibshirani, R. J. 1993. *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Everett, M. D. 1977. Student evaluations of teaching and cognitive level of economics courses. *Journal of Economic Education*. 8: 100-103.
- Felton, J., Koper, P.T., Mitchell, J. and Stinson, M. 2008. Attractiveness, easiness and other issues: student evaluations of professors on *ratemyprofessors.com*. *Assessment and Evaluation in Higher Education*. 33: 45-61.
- Greene, W. H. 2000. *Econometric Analysis*. New Jersey: Prentice-Hall.
- McKelvey, R. D. and Zavoina, W. 1975. A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*. 4: 103-20.
- Selvanathan, A., Selvanathan S., Keller and Warrack, B. 2006. *Australian Business Statistics* (4<sup>th</sup> Edition), Melbourne: Thomson Learning.
- Snee, R.D. 1973. Some aspects of nonorthogonal data analysis, Part 1. Developing prediction equations. *Journal of Quality Technology*. 5, 67-79.
- Tang, T. L-P. 1997. teaching evaluation at a public institution of higher education: factors related to overall teaching effectiveness. *Public Personnel Management*. 26: 379-89.
- Wilson, R. 1998. New Research casts doubt on value of student evaluations of professors. *Chronicle of Higher Education*. (January 16): A12.

## **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 Arthashastra: 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 'Arthashastra' 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.
42. Students' Evaluation of Teaching Effectiveness: What Surveys Tell and What They Do Not Tell by Clem Tisdell and Mohammad Alauddin, November 2006.
43. Economic Prospects for Small Island Economies, Particularly in the South Pacific, in a Globalising World by Clem Tisdell, November 2006.
44. The Evolution and Classification of the Published Books of Clem Tisdell: A Brief Overview by Clem Tisdell, July 2007.
45. Cost-Benefit Analysis of Economic Globalization by Clem Tisdell, January 2008.
46. Economic Benefits and Drawbacks of Cities and their Growth Implications by Clem Tisdell, January, 2008.
47. Interfirm Networks in the Indonesian Garment Industry: Trust and Other Factors in their Formation and Duration and their Marketing Consequences by Latif Adam and Clem Tisdell, April, 2008.

48. Trust and its Implications for Economic Activity, Welfare and Globalisation by Clem Tisdell, April, 2008.
49. Economics, Corporate Sustainability and Social Responsibility by Clem Tisdell, May 2008.
50. Structural Transformation in the Pig Sector in an Adjusting Vietnam Market: A Preliminary Investigation of Supply-side Changes by Clem Tisdell, September 2008
51. Thirty Years of Economic Reform and Openness in China: Retrospect and Prospect by Clem Tisdell, October 2008.