ECONOMIC ANALYSIS & POLICY, VOL. 40 NO. 3, DECEMBER 2010

Do Instructional Attributes pose Multicollinearity Problems? An Empirical Exploration

*Mohammad Alauddin**
School of Economics, University of Queensland,
Brisbane, QLD 4072, Australia
(Email: m.alauddin@uq.edu.au)

and

Hong Son Nghiem
Centre of National Research on Disability & Rehabilitation Medicine,
School of Medicine, University of Queensland,
Brisbane, QLD 4072, Australia

Abstract: It is commonly perceived that variables 'measuring' different dimensions of teaching (construed as instructional attributes) used in student evaluation of teaching (SET) questionnaires are so highly correlated that they pose a serious multicollinearity problem for quantitative analysis including regression analysis. Using nearly 12000 individual student responses to SET questionnaires and ten key dimensions of teaching and 25 courses at various undergraduate and postgraduate levels for multiple years at a large Australian university, this paper investigates whether this is indeed the case and if so under what circumstances. This paper tests this proposition first by examining variance inflation factors (VIFs), across courses, levels and over time using individual responses; and secondly by using class averages. In the first instance, the paper finds no sustainable evidence of multicollinearity. While, there were one or two isolated cases of VIFs marginally exceeding the conservative threshold of 5, in no cases did the VIFs for any of the instructional attributes come anywhere close to the high threshold value of 10. In the second instance, however, the paper finds that the attributes are highly correlated as all the VIFs exceed 10. These findings have two implications: (a) given the ordinal nature of the data ordered probit analysis using individual student responses can be employed to quantify the impact of instructional attributes on TEVAL score; (b) Data based on class averages cannot be used for probit analysis. An illustrative exercise using level 2 undergraduate courses data suggests higher TEVAL scores depend first and foremost on improving explanation, presentation, and organization of lecture materials.

* Corresponding author Mohammad Alauddin.

I. INTRODUCTION

The practice of student evaluation of teaching (SET) has experienced a phenomenal growth over the last three decades or so (Seldin 1998). University administrators invariably use SET as important indicator of an instructor's pedagogical contribution to teaching and learning and hence teaching effectiveness.

Much criticism of and support for Student Evaluation of Instruction (TEVAL) scores as a measure of teaching effectiveness (see for example, Felton et al. 2008; Bursdal and Harrison 2008), notwithstanding detailed quantitative measurement of the impacts of instructional attributes used in SET on TEVAL scores are lacking (Alauddin & Tisdell 2010).

Furthermore, there is a common perception that variables 'measuring' different dimensions of teaching (construed as instructional attributes) used in SET questionnaires are so highly correlated that they pose a serious multicollinearity problem (DeBerg and Wilson 1990, Martin 1998, Kulik 2001) which is often regarded as an impediment to quantitative analysis such as ordered probit models. In light of the above, the central idea canvassed in the paper is that while one cannot altogether rule out the multicollinearity problem, in reality it may not be so serious as to render measurement of quantitative impact of instructional attributes on TEVAL scores useless. Using a large data set of individual student responses to the SET questionnaires with ten key dimensions of teaching at a large Australian university, this paper investigates whether the extent to which this might be the case. Note that this is an econometric exercise and does not address any theoretical issues underlying the teaching and learning process in higher education. The paper is organised as follows. Section 2 provides a brief overview of the data. Section 3 presents methodology and empirical results, and provides a discussion of results. Section 4 presents some results of ordered probit analysis as an illustrative example. Section 5 presents conclusions.

II. THE DATA

The basic data for this study are from the SET surveys for economics courses at the University of Queensland. These relate to the period 2000-2007 and are based on nearly 12000 completed SET forms involving 25 courses and 102 student cohorts and 20 lecturers. The courses included eighteen undergraduate level courses (5 level 1, 6 level 2 and 7 level 3) and 7 postgraduate courses.

The data do not meet the criterion of strict randomness as the courses 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 wide range of courses – including large-sized, first, second and third level undergraduate and postgraduate courses. These courses display considerable diversity in their student populations typified, amongst other things, by academic background, study discipline, sex, ethnicity, and student quality. Note that the University of Queensland requires all instructors to collect TEVAL data. 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. Particularly, students who are less likely to attend classes are under-represented since 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*), and electronic communications and so on. Nevertheless, those students attend and determine an instructor's TEVAL score.

Table 1 provides the codes and definitions of dependent and independent variables used in this study.

Table 1: Definitions and Description of Instructional Attributes used in SET Data

Variable Code	Description
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: Instructional attributes (1 – strongly disagree, 5 – strongly agree)	
ORGANIZE	The lecturer produced classes that were well organized
PRESENT	The lecturer presented material in an interesting way
FEEDBACK	The lecturer gave adequate feedback on my work
RESPECT	The lecturer treated students with respect
KNOWWELL	The lecturer seemed to know the subject well
ENTHUSM	The lecturer communicated her/his enthusiasm for the subject
THINKMEM	The lecturer emphasized thinking rather than memorizing
EXPLAIN	The lecturer gave explanations that were clear
CONSULT	The lecturer was available for consultation
LSKILLS	The lecturer helped to improve my learning skills

III. METHODOLOGY, RESULTS AND DISCUSSION

Originally introduced by Frisch (1934), strictly speaking, "*multicollinearity* refers to the existence of more than one exact relationship, and *collinearity* refers to the existence of a single linear relationship. But this distinction is rarely maintained in practice, and multicollinearity refers to both" (Gujarati 2003, p.342). The variance inflation (inflating) factor (VIF) is often used to test the extent of multicollinearity. It is defined as:

$$\mathrm{VIF}_k = \frac{1}{1-R_k^2}$$

where R_k^2 is the R^2 in the regression of x_k on all the other independent variables (Greene, 2003. p.57). VIF shows how the presence of multicollinearity inflates the variance of an estimator. As R_k^2 approaches 1, VIF approaches infinity. In the absence of any multicollinearity, R_k^2 will be close to zero and VIF will approach unity. The inverse of VIF is called tolerance (TOL). If VIF_k is very high TOL will be very low. In the literature VIF and TOL are used interchangeably

(Gujarati 2003, p.353).

As a rule of thumb, multicollinearity may not be a serious issue if VIF does not exceed 10 (i.e. TOL > 0.1), although some authors use a more conservative rule that VIF does not exceed 5. However, O'Brien (2007) suggests that this rule of thumb should be assessed in a contextual basis, taking into account factors that influence the variance of regression coefficients. O'Brien argued that the VIF value of 10 or even 40 or higher does not suggest the need for common treatment of multicollinearity such as using ridge regressions, elimination of some variables, or combine them into a single index. Despite the criticism against VIF it is very widely as tool of detecting multicollinearity. This paper uses VIF as an indicator of multicollinearity while acknowledging its limitations.

3.1 VIF Values for Aggregate Data- Class Averages

Table 2 presents the VIF values for all the instructional attributes with class averages. It is clear that at the class average level the attributes are highly correlated, as demonstrated by the very high VIFs which range between 66.42 for LSKILLS and 8.65 for ORGANISE. Given serious multicollinearity, regression analysis cannot be used to data at the class-average level.

Table 2: Variance Inflation Factors for Class Average Characteristics

Variables	VIF
LSKILLS	66.43
PRESENT	35.10
KNOWWELL	26.82
EXPLAIN	26.46
ENTHUSM	25.20
THINKMEM	20.25
FEEDBACK	18.68
RESPECT	18.40
CONSULT	17.90
ORGANISE	8.65

3.2 VIF Values for Aggregate Data- Individual Responses

1. Aggregated for All and Individual Years

Table 3 presents the VIF results for aggregate data for all years and annually. Because of the missing values, the valid number of observations ultimately stood at 11634 student responses. The information contained in *Table 3* suggests that, using a lenient rule of thumb that VIF 5, multicollinearity is not a problem for results of neither for all years nor for the individual years. For all years, only one VIF exceeds 2.50. However, when VIF values are considered for individual years, some values exceed this threshold. For example, for 2000, 7 out of 10 attributes, the VIF values exceed 2.50 whilst in 2002, 2003 and 2007 respectively 4, 3 and 8 VIF values exceed this threshold.

Table 3: Variance Inflation Factors for All Courses by Year

Variables	All years	2000	2001	2002	2003	2004	2005	2006	2007
PRESENT	2.51	3.33	2.25	3.17	2.71	2.41	2.40	2.22	2.58
ENTHUSM	2.47	2.95	2.28	2.74	2.40	2.43	2.47	2.20	3.63
LSKILLS	2.25	2.72	2.25	2.60	2.44	2.15	2.08	1.98	2.68
THINKMEM	2.24	2.71	2.14	2.37	2.20	2.22	2.18	2.05	2.58
KNOWWELL	2.19	2.57	2.14	2.43	2.61	2.29	2.13	1.96	3.66
ORGANISE	2.08	3.24	1.97	2.40	2.28	1.94	1.86	1.96	2.47
EXPLAIN	1.91	3.62	2.15	2.97	2.61	2.31	2.20	1.29	2.91
FEEDBACK	1.77	1.78	1.80	1.82	1.86	1.67	1.74	1.69	2.15
CONSULT	1.61	1.43	1.65	1.58	1.57	1.58	1.70	1.59	2.77
RESPECT	1.51	2.17	1.89	1.45	2.13	2.01	1.85	1.15	3.27
N	11634	759	1554	1201	1318	1670	2218	2668	246

2. Aggregated All Years but for Separate Levels

Table 4 provides VIF values based on data disaggregated by levels. It can be found that only in nine out of fifty VIF values, can one find VIF greater than 2.50 and none exceeds a value of 3. In general, for postgraduate courses, the VIF values are higher than those for any of the undergraduate courses. For six out of ten attributes, the VIF values exceed 2.50.

Table 4: Variance Inflation Factors by Course Level for All Years Combined

Variables	Level 1 undergraduate	Level 2 undergraduate	Level 3 undergraduate	Postgraduate
ENTHUSM	2.70	2.21	2.10	2.96
PRESENT	2.67	2.42	2.48	2.55
THINKMEM	2.34	2.00	2.26	2.58
LSKILLS	2.27	2.09	2.11	2.86
KNOWWELL	2.13	2.39	2.04	2.71
ORGANISE	2.06	2.14	2.27	2.30
FEEDBACK	1.74	1.66	1.80	1.97
EXPLAIN	1.66	2.54	2.70	2.55
CONSULT	1.60	1.47	1.62	1.89
RESPECT	1.40	1.90	1.35	2.14
N	5868	2805	1511	1450

While the results presented in *Tables 3* and *4* provide some indications on the extent of multicollinearity, the VIF values are estimated based on higher level of aggregation. For example, in *Table 3*, no distinction is made between level 1 and level 3 undergraduate student responses or between undergraduate and postgraduate responses. It is useful to consider the VIF values in case of more disaggregated data which we undertake in Section 3.2.

3.3 VIF Values for Disaggregated Data

1. Level 1 Undergraduate Non-Quantitative Course

Table 5 presents VIF values for Level 1 undergraduate non-quantitative courses. As can be seen only nine of the seventy VIF values exceed 2.50. These relate to two values for all years, two in 2003, three in 2004 and two in 2005. None of the VIF values exceed three.

Table 5: Variance Inflation Factors for Level 1 Undergraduate Principles Course (Non-Quantitative) with Various Lecturers by Year

Variables	All years	2000	2001	2002	2003	2004	2005	2006
PRESENT	2.67	8.40	2.14	3.71	2.46	2.39	2.66	2.14
ENTHUSM	2.70	4.92	2.18	3.30	2.46	2.65	2.60	2.25
LSKILLS	2.27	7.23	2.18	2.88	2.16	2.10	2.15	1.91
THINKMEM	2.34	4.73	2.24	3.15	2.03	2.20	2.15	1.97
KNOWWELL	2.13	3.76	2.14	2.47	2.60	2.30	2.03	1.98
ORGANISE	2.06	5.90	2.01	3.07	2.08	1.90	1.75	1.83
EXPLAIN	1.66	8.73	2.22	3.82	2.58	2.29	2.17	1.11
FEEDBACK	1.74	3.15	1.57	1.89	1.76	1.55	1.68	1.67
CONSULT	1.60	2.28	1.46	1.60	1.64	1.48	1.73	1.64
RESPECT	1.40	2.57	1.78	2.12	2.19	2.05	1.88	1.10
N	5868	178	760	325	435	1159	1397	1614

2. Level 1 Undergraduate Quantitative Course

Referring to *Table 6*, in one case (2000) the VIF exceeds 5. However, in the same year, seven other VIF values exceed 2.5. For all years, six out of ten VIF values exceed 3.5. In subsequent years (2001 and 2002) four out of ten in each case exceed 2.50. However, for 2000 four VIF vales exceed 3.50. Overall, except for one case that VIF exceed 5, there was no evidence of serious multi-collinearity in this category.

Note, however, that the VIF values seem to be somewhat higher for the quantitative courses than for the non-quantitative causes. It would have been useful to provide VIF values by lecturers. However, we did not have sufficient data for each lecturer to provide a meaningful basis of comparison for Level 1 courses.

Table 6: Variance Inflation Factors for Level 2 Undergraduate (Quantitative) Courses with Various Lecturers by Year

Variables	All years	2000	2001	2002	2003	2004	2005	2006
PRESENT	2.42	2.12	1.93	2.91	3.27	2.03	2.10	2.41
ENTHUSM	2.21	1.87	2.09	2.63	2.54	1.68	2.04	2.45
LSKILLS	2.09	1.79	1.92	2.45	2.87	1.86	1.78	1.81
THINKMEM	2.00	2.04	1.93	2.01	2.50	1.76	1.94	1.82
KNOWWELL	2.39	1.90	2.17	3.03	2.96	1.95	2.31	2.45
ORGANISE	2.14	1.97	1.92	2.57	2.88	1.76	1.80	1.82
EXPLAIN	2.54	2.24	2.02	3.15	3.44	2.31	2.08	2.05
FEEDBACK	1.66	1.44	1.71	1.71	2.01	1.72	1.77	1.83
CONSULT	1.47	1.26	1.56	1.34	1.50	1.52	1.76	1.80
RESPECT	1.90	1.68	1.76	1.92	2.38	1.77	1.67	2.03
N	2,805	403	407	404	485	288	432	386

3.4 Same Lecturer Teaching Two Courses at Two Levels

We were able to gather a consistent data series for each of seven years for two courses taught by the same lecturer. One course is a second-level undergraduate non-quantitative course while the other is an introductory postgraduate quantitative course. The student populations in the undergraduate course are primarily from the English-speaking background with Australian schooling and are enrolled in the economics degree program (single or double major). On the other hand, the student populations in the postgraduate course are predominantly from non-English speaking background (NESB) with schooling outside of Australia and with at least 80 per cent enrolled in non-economics major especially commerce. A characteristic feature of the populations in the postgraduate course is the high incidence of statistics anxiety (Alauddin and Butler 2004; Onwuegbuzie 2000).

Table 7 presents VIF values for the instructional attributes for the level 2 undergraduate course for seven years and for all years combined. In twenty-nine out of 80 cases do VIF values exceed 2.5, including all the ten VIF values in 2005. However, none of these VIF values comes anywhere close to the lower threshold value of 5.

Table 8 sets out VIF values for the postgraduate courses. In seven of the eighty cases, VIF values exceed 5 (three in 2001 and four in 2004). Furthermore, nine VIF values exceed the conservative threshold of 5 In general, the VIF values for the postgraduate course are higher than those for the undergraduate course given the same lecturer.

Why this might be so? It is plausible that differences in English language competency of majority of students from undergraduate (predominated by native English speaking (ESB) students) and postgraduate (predominated by non-English speaking (NESB) students) courses might make some differences in the way students from two distinct populations interpret the SET questionnaire. An NESB student may not interpret the survey questionnaires correctly or look at one or two attributes (e.g. presentation and enthusiasm) and rate other instructional attributes accordingly. This might suggest the presence of what in the educational literature is known as halo effect (Engdahl et al. 1993) which refers to a cognitive bias whereby the perception of a particular attribute influences the perception of other attributes. For example, a student's perception of an instructor might be highly influenced by her/his presentation of lecturer materials.

Table 7: Variance Inflation Factors for a Level 3 Undergraduate (Non-Quantitative) Course with Same Lecturer by Year

Variables	All years	2000	2001	2002	2003	2004	2005	2006	2007
PRESENT	2.48	2.52	2.48	2.53	2.40	3.72	2.32	2.68	2.67
ENTHUSM	2.10	2.72	2.35	1.93	2.31	2.66	2.14	2.09	3.60
LSKILLS	2.11	2.70	2.08	1.77	2.09	3.24	2.22	2.60	2.40
THINKMEM	2.26	2.01	2.43	1.85	2.60	2.91	2.64	2.28	2.86
KNOWWELL	2.04	2.50	2.33	2.09	2.20	2.95	1.64	2.04	4.85
ORGANISE	2.27	2.02	1.97	2.57	2.30	2.67	2.29	2.99	2.39
EXPLAIN	2.70	2.36	2.32	2.58	2.51	4.30	2.64	3.74	2.88
FEEDBACK	1.80	3.22	1.89	1.76	1.78	2.25	1.72	2.50	1.81
CONSULT	1.62	1.69	1.99	1.65	1.56	2.06	1.87	1.65	2.67
RESPECT	1.35	2.23	2.19	1.05	1.83	2.41	1.91	2.07	3.10
N	1,511	178	296	228	259	159	124	205	62

Table 8: Variance Inflation Factors for an Introductory Postgraduate Statistics Course with the Same Lecturer by Year

Variables	All years	2001	2002	2003	2004	2005	2006	2007
PRESENT	2.55	2.46	2.68	2.33	4.64	2.61	2.46	2.75
ENTHUSM	2.96	6.39	2.70	2.14	5.01	4.25	2.34	4.04
LSKILLS	2.86	5.26	3.43	2.05	5.43	2.97	2.76	3.13
THINKMEM	2.58	6.06	2.64	2.21	4.58	3.20	2.40	2.69
KNOWWELL	2.71	4.24	2.72	2.97	5.19	2.94	2.08	3.70
ORGANISE	2.30	4.55	2.33	2.18	4.00	2.89	2.14	2.57
EXPLAIN	2.55	2.06	2.21	1.42	6.54	2.91	2.80	3.36
FEEDBACK	1.97	3.23	2.39	2.07	3.87	2.05	1.83	2.54
CONSULT	1.89	4.28	2.57	2.07	2.66	1.67	1.64	2.97
RESPECT	2.14	3.99	2.36	1.94	2.72	2.36	1.77	3.54
N	1,450	91	244	139	64	265	463	184

It is clear from the above that despite instructional attributes being correlated, the extent of correlation is unlikely to render the regression results redundant when individual student responses are used. The next section provides an illustrative exercise.

IV. AN ILLUSTRATIVE EXAMPLE: ORDERED PROBIT ANALYSIS

One important implication of the results stemming from this research is that regression analysis can be employed to quantify the impact of the instructional attributes on TEVAL score. However, a large body of literature recognises 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). Greene (2003, p.736) 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 2003). Since McKelvey and Zovoina (1975), these models have been widely used as a methodological framework for analysing ordered data.

Table 9: Ordered Probit Analysis of Quantitative Impacts of Instructional Attributes TEVAL for Level 2 Economics Courses: An illustrative Example.

Variables	Coefficient	Standard error	Marginal effects		
			P(TEVAL=3)	P(TEVAL=4)	P(TEVAL=5)
			(9%)	(74%)	(16.8%)
ORGANISE	***0.36	0.04	***-0.06	***-0.03	***0.09
PRESENT	***0.49	0.04	***-0.08	***-0.04	***0.12
FEEDBACK	0.02	0.03	-0.003	-0.001	0.004
RESPECT	***0.20	0.04	***-0.03	***-0.02	***0.05
KNOWWELL	***0.18	0.05	***-0.03	***-0.02	***0.05
ENTHUSM	***0.17	0.04	***-0.03	***-0.01	***0.04
THINKMEM	***0.23	0.04	***-0.04	***-0.02	***0.06
EXPLAIN	***0.48	0.04	***-0.08	***-0.04	***0.12
CONSULT	*0.05	0.03	*-0.01	*-0.005	*0.01
LSKILLS	***0.28	0.04	***-0.05	***-0.02	***0.07
μ_1	***4.51	0.21			
μ_2	***6.10	0.21			
μ_3	***8.46	0.25			
μ_4	1***0.75	0.28			

Pseudo-R^2: 0.46 , N=2529, LR $\chi^2(20)$=2,868 p < .0001; ***, ** and * represent the significant levels of 1%, 5% and 10%, respectively.

For example, an ordered probit analysis of level 2 course reveals that, with the exception of FEEDBACK, all instructional attributes significantly affect the course TEVAL. Particularly, if a lecturer increase his/her EXPLAIN score by one unit from the median value of 4, s/he will have the probability of having a TEVAL=5 increased by 15.1 per cent whilst the probability of having a TEVAL of 3 and 4 decreased by 5.4 per cent and 9.77 per cent, respectively (see *Table 9*). Note that the probability of having a TEVAL score of 1 and 2 is basically zero at 3 decimal places, hence they are not reported for brevity. PRESENT and ORGANISE are the next two important attribute of lecturing as one unit increase from the median value would lead to an increase on the probability of having a TEVAL=5 by 15.4 and 11.1 per cent, respectively. LSKILLS is also a very important attribute of lecturers and its marginal effects estimate at the median for the probability of having a TEVAL=5 is 8.8 per cent. The remaining attributes such as RESPECT, KNOWWELL and ENTHUSM all significantly affect the TEVAL of lecturing but their magnitudes are slightly smaller than the above mentioned attributes. In short, according to students' evaluation in second level economics courses, improvement in explanation, presentation, and organization of lecture materials play important roles in their rating of the lecturer.

V. CONCLUSIONS

An examination of the variance inflation factors (VIFs), across courses, levels and over time, finds no sustainable evidence of multicollinearity. While there were a few isolated cases of VIFs marginally exceeding the conservative threshold of 5, in no cases did the VIFs for any of the instructional attributes come anywhere close to the high threshold value of 10. The findings of this paper, therefore, cast doubt about the validity of the claim of any serious multicollinearity problem involving instructional attributes.

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 masks it. The application of ordered probit analysis to a subset of the data (second-level undergraduate courses) reveals that an instructor may need to focus on improving presentation, explanation and organization of lecture materials.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge constructive comments and suggestions by the editor and an anonymous referee of this journal. An earlier version of this paper was presented at the Oxford Business & Economics Conference, St. Hugh's College, Oxford, June 2009. The authors wish to thank the discussant and participants at the conference for useful comments. Usual *caveats* apply.

REFERENCES

Alauddin, M. and J.E. Butler (2004). 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: 202-08.

Bursdal, C.A. and P.D. Harrison (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.

Engdahl, R.A., R.J. Keating, and J. Perrachoine (1993). Effects of grade feedback on student evaluation of instruction, *Journal of Management Education*. 17**:** 174-84.

DeBerg, C.L. and J.R. Wilson (1990). An empirical investigation of the potential confounding variables in student evaluation of teaching, *Journal of Accounting Education*. 8: 37-63.

Felton, J., P.T. Koper, J. Mitchell, and M. Stinson (2008). Attractiveness, easiness and other issues: Student evaluations of professors on ratemyprofessors.com, *Assessment and Evaluation in Higher Education*. 33: 45-61.

Frisch, R. (1934). Statistical Confluence Analysis by Means of Complete Regression Systems, Oslo University, Institute of Economics, publication no. 5.

Greene, W. (2003). *Econometric Analysis* (5th Edition). Upper Saddle River: Prentice-Hall.

Gujarati, D. (2003). *Basic Econometrics*. New York: McGraw-Hill.

Kulik, J.A. (2001). Student ratings: Validity, utility, and controversy, in: M. Theall, P.C. Abrami, and L.A. Mets (eds.), *New directions for institutional research: No. 109. The student ratings debate: Are they valid? How can we best use them*?. San Francisco: Jossey-Bass: 9–26.

McKelvey, R.D. And W. Zavoina (1975). A statistical model for the analysis of ordinal level dependent variables, *Journal of Mathematical Sociology*. 4: 103-20.

Martin, J.R. (1998). Evaluating faculty based on student opinions: Problems, implications and recommends from Deming's theory of management perspective, *Issues in Accounting Education*. 13: 1079–1094.

Marquardt, D.W. and R.D. Snee (1974). Test Statistics for Mixture Models, *Technometrics*. 16: 533-37.

O'Brien, R.M. (2007). A caution regarding rule of thumb for variance inflation factors, *Quality and Quantity*. 41: 673-90.

Onwuegbuzie, A.J. (2000). Statistics anxiety and the role of self-perception, *Journal of Educational Research*. 63: 323-30.

Snee, R.D. (1973).Techniques for analysis of mixture data, *Technometrics*. 15: 517-28.