THE CORRELATION BETWEEN TEACHING ATTRIBUTES AND THE INSTRUCTOR'S RATING

Luther D. Lawson, UNC at Wilmington

ABSTRACT

This paper examines the relationship between teaching effectiveness and the awarding of economic and finance grades at one large regional institution. Using several statistical tests, the paper compared the average course grade that the economics and finance faculty members gave their students and the average score that the students awarded their instructor on the university's teaching effective instrument. The findings suggest that for the sample groups, there was no correlation between the class grade and the teacher's "grade."

INTRODUCTION

At the conclusion of most academic semester, it has become a common practice at many colleges and universities to ask for student feedback concerning the instructor's effectiveness. Ostensibly, the purpose is to provide the teacher with information on their strengths and weaknesses in order that areas needed for improvement can be identified and acted upon. The number of instruments used to measure teaching effectiveness and various attributes of the instructor or the course, however, are as varied as there are colleges and universities. Most instruments of measurement are grouped by pedagogy and other categories designed to capture specific attributes relative to the instructor. Regardless of the format of the assessment instrument used, and the several questions found on that instrument, many evaluative instruments contain but one question that is designed to encapsulate the instructor's *overall* effectiveness for that particular course. What is often overlooked is the relevance of the other questions as predictors of the instructor's teaching effectiveness.

While teaching evaluations can provide valuable information for the instructor of record, many faculty members, however, feel that some department chairs might rely on poor evaluations as a means for not granting promotion, tenure, or merit raise. If there is merit to the latter argument then some faculty members may "buy" good teaching evaluations by awarding high and unearned grades. In casual conversations, some professors who have received poor evaluations will counter by claiming that the reason lies in their tough grading policy and rigorous standards. These same faculty members will claim that the way to insure high teaching marks is to award inflated grades. While this paper lays no claim to provide definitive answers to the on-going controversy, for there are multiple reasons, which might explain high grades, this inquiry could shed some preliminary light. In this regard, the purpose of the paper is to examine the correlation between measures of teaching effectiveness and class grades. The null hypothesis is that there is a positive relationship between teaching effectiveness and class grades, due to one "buying" grades.

DATA SET AND STATISTICAL TESTS

The data sets are from the economics and finance department of a mid-size southeastern university. The university requires that all faculty members allow their students to evaluate the course, instructor, etc., at the end of each semester. The instrument used to "measure teaching effectiveness", Student Perception of Teaching (SPOT), is composed of 16 questions, according to Pedagogy, Personal Characteristics, Ethical and Personal Characteristics and Overall Rating. Each question is designed to capture some aspect of the course or the instructor of record, and allows the student to respond with one of five responses: poor, below average, average, above average, or excellent. The weight assigned to each response ranges from 1 (poor) to 5 (above average). One question in particular, 'Instructor Rating Compared to All Others' (Q16), has become *the* question that university administrators hold to be the "best" indicator of teaching effectiveness. The numerical response to this question shall serve as one critical variable of teaching effectiveness, in this paper and his coded as EFFECTIVE. The other critical variable will be coded as GRADE. Other quantitative variables found on the SPOT are coded as Q1-Q16 and are described as:

Further, at the culmination of each semester the chair is furnished with the number of A's, B's, C's, etc. that the faculty awarded for each class that he or she taught. Given that an A=4 pts., and a B=3 pts., etc., the average grade for the class can be determined. Hence, GRADE, became the other critical variable in the analysis.

The data set consisted of all tenured or tenured track faculty members in the department of economics and finance (25 professors) and covered a period of five academic years including the spring semester 2001 (1996-2001). It did not include summer sessions for this period. Further, where either GRADE or EFFECTIVE was not available, that particular class was eliminated from the data set. In all there were

325 observations. Lastly, other variables were identified and will be used in several different models. In particular: Classize, Core, semester, and tenured vs. non-tenured. All will be described below.

STUDENT PERCEPTION OF TEACHING (SPOT)		
Pedagogy -		
	Presentation of Material	= student's score on Q1
	Learning Experience	= student's score on Q2
	Organization of Course	= student's score on Q4
	Work required of Course	= student's score on Q5
	Explanation of Assignments	= student's score on Q7
	Relevance of Assignments	= student's score on Q8
	Feedback on Assignments	= student's score on Q9
	Instructor's Expectations	= student's score on Q6
Personal Characteristics –		
	Stimulation of Interest	= student's score on Q3
	Concern for Students	= student's score on Q10
	Accessibility of Instructor	= student's score on Q11
	Instructor's Enthusiasm	= student's score on Q14
Ethical and Intellectual Characteristics –		
	Fairness in Grading	= student's score on Q12
	Knowledge of Subject	= student's score on Q13
Overall –		
	Instructor Rating Compared	=student's score on Q16 (EFFECTIVE)
	Course Rating Compared	=student's score on Q15
[Question 16 serves as one of two key dependent variables and is reported as EFFECTIVE. The other key dependent variable is reported as GRADE. Both will be defined below.]		

STATISTICAL TESTS AND MODELING

Two null hypotheses will be tested: there is no statistical and positive relationship between the high GRADE and high student evaluations (EFFECTIVE) and there is no statistical relationship between GRADE and SPOT scores. To accept or reject the hypothesis several statistical tests were performed.

First, an average of EFFECTIVE scores (Q16) was ranked from high to low for all members of the department. The purpose was to determine statistically significance from the department's average, by all professors in the department of economics and finance. Table 1 records the scores from high to low. Professors A-E was found to be statistically significant from professors F-T. Professors U-Y was likewise found to be statistically significant from professors F-T. Second, a Pearson Correlation Coefficient Matrix (not shown) was constructed for all questions found on the teaching evaluation instrument (SPOT) numbered 1-16. The purpose was to ascertain any correlation among any questions on this instrument. The Pearson Correlation matrix revealed that questions numbered 5 (Work required), 6 (Instructor" Expectation), 9 (Feedback on Assignments) and 15 (Course Comparison) were highly correlated, each were omitted from the models. Third, a series of OLS models were constructed where the dependent variable EFFECTIVE or GRADE was regressed on variables GRADE, CBK, SEMESTER, TENURE and questions 1-14.

The variables are specified as in the OLS models are as follow. Tables 2-5 records the results found for these models.

GRADE	= the average grade given by the instructor
SEMESTER	= dummy variable: 1= fall semester, 0= spring semester
PROF	= Professors in the data set
CBK	= dummy variable: 1= core elective course, 0= required core course CBK (Common Body of Knowledge, are those courses that all students in the business school must take, regardless of major)
TENURED	= dummy variable: 1 = nontenured status, 0 = tenured status
CLASSIZE	= the enrollment of the class, where: *small* classes are less than 13 & *large* classes are greater than 12

STATISTICAL RESULTS

In order to identify those professors that statistically receive EFFECTIVE scores above and below the mean, a Least Squares Mean test was conducted. Table 1 records the results. Basically, Table 1 is a one-way ANOVA model, which explains EFFECTIVE by professor, controlling for no other variables. As seen in columns 1 and 3, five departments member's EFFECTIVE scores were statistically significant *above* the department's average, and five were statistically significant *below* the departmental average. Based on the fact that twenty-five professors make up the data pool, it was decided to focus attention on the top five faculty members (A-E) and bottom five faculty members (U-Y) to determine if *these* instructors "buy" high teaching evaluations. Table 2 records the relationship between four predictor variables and the dependent variable, EFFECTIVE. What clearly emerges from the data is that for the top faculty, there is no relationship between EFFECTIVE and GRADE. Interestingly, no other variables for this group prove to be a factor in "predicting" the instructor's effectiveness. For the bottom faculty, the interpretation of the results depends on the reader. The correlation between GRADE and EFFECTIVE are statistically and positively related. Recognizing that this group is the ones that score the lowest EFFECTIVE score, then as their scores fall, so do the grades. Equally plausible, however, is that as their EFFECTIVE scores rise so do the grades.

Attention is then given to all department members, regardless of their EFFECTIVE score ranking, but based on class size. It was hypothesized that small classes (less than 13) would yield different results than classes greater than 13. Table 3 sought to determine if there was any "predictive" value between EFFECTIVE, GRADE, CBK, SEMESTER and TENURE, by large and small classes. The table is set in the aggregate and does not account for individual faculty members. With respect to EFFECTIVE and GRADE, the table does show a positive and statistically significant relationship, regardless of the class size. The data would appear to support the argument for "buying teaching evaluations." Table 3 reveals that no other variable are statistically significant. Clearly, more analysis is needed.

Tables 4 and 5 record the regression of SPOT against GRADE and are segmented by tenured and non-tenured faculty. To better interpret the data, the results the two tables should be juxtaposed. The results indicate that there are several teaching attributes that are statistically common for both large classes and small classes, regardless of tenure status. For both tenure sets, only Learning Experience (2) and Explanation of Assignments (Q 7) as a factor in determining grade. Interestingly, the variable CBK is non-significant (-1.39) for small classes and significant at –2.61 for tenured faculty. For large classes the results are mixed. In the area of *Pedagogy,* only Presentation of Material (Q1) emerges as important, regardless of tenure status. However, in the area of *Personal Characteristics*, both Stimulation of Interest (Q3) and Concern for Students (Q10) are consistently significant, regardless of tenure status.

Further, once again, the variable CBK emerges as a "predictive" factor, but of a curious sign.

Table 1: Mean Score Ranking of All Professors: Effective

Professor	Mean Effective Score	Statistical Significance (5%)
A	4.74	*
B	4.56	*
C	4.5	*
D	4.48	*
E	4.43	*
F	4.3	
G	4.25	
H	4.16	
I	4.14	
J	4.12	
K	4.09	
L	4.09	
M	4.07	
N	4.04	
O	4	
P	3.98	
Q	3.88	
R	3.81	
S	3.71	
T	3.66	
U	3.61	*
V	3.58	*
W	3.54	*
X	3.43	*
Y	3.03	*

Table 2: Relationship Between Predictor Variables and Effective: Select Faculty			
(Top Faculty)			
Variables	Parameter Estimate	t Value	PR >|T|
INTERCEPT	4.6945	13.41	0.0001
GRADE	-0.1704	-0.17	0.8683
CBK	-0.1302	-1.42	0.1615
SEMESTER	0.0328	0.55	0.5817
TENURE	-0.0718	-1.13	0.2634
R^2: 070; n: 66; F: 1.11			
(Bottom Faculty)			
INTERCEPT	1.3502	3.69	0.0005
GRADE	0.7513	6.15	0.0001
CBK	0.0149	0.08	0.9335
SEMESTER	-0.1095	-1.79	0.0778
TENURE	0.4472	2.88	0.0057
R^2: 40: n: 69: F: 10.92			

Table 3: Relationship Between Predictor Variables and Effective All Faculty			
(For Class Size ≤ 12)			
Variables	**Parameter Estimate**	**t Value**	**PR >|T|**
INTERCEPT	2.834	5.63	<.0001
GRADE	0.634	4.79	<.0001
CBK	0.234	1.81	0.0742
SEMESTER	-0.039	-0.36	0.7208
TENURE	-0.0695	-0.64	0.5249
R^2: .2197; n: 86; F: 5.77			
(For Class Size > 12)			
INTERCEPT	2.8398	10.48	<.0001
GRADE	0.4621	5.31	<.0001
CBK	-0.0246	-0.36	0.7208
SEMESTER	-0.0638	-1.09	0.2767
TENURE	-0.0226	-0.37	0.7135
R^2: 1483: n: 253: F: 10.84			

Table 4: Relationship Between Predictor Variables and GRADE: Non-Tenured Faculty			
(Class Size ≤ 12)			
Variables	**Parameter Estimate**	**t Value**	**PR >|T|**
INTERCEPT	2.3762	3.49	0.0016
Pedagogy			
(Q2) Learning Experience	0.2283	0.83	0.4123
(Q7) Explanation of Assignment	-0.0489	-0.16	0.8716
Ethical & Intellectual			
(Q13) Knowledge of Subject	-0.0248	-0.12	0.9072
CBK	-0.2283	-1.39	0.1754
R^2: .10; n: 33; F: .81			
(For Class Size > 12)			
INTERCEPT	2.9327	8.61	0.0001
Pedagogy			
(Q1) Presentation of Material	-0.3439	-2.16	0.0326
(Q4) Organization of Course	-0.2921	-2.2	0.0298
Personal Characteristics			
(Q3) Stimulation of Interest	0.3191	2.6	0.0106
(Q10) Concern for Students	0.3388	3.17	0.002
(Q11) Accessibility of Instructor	0.0017	2.8	0.006
CBK	-0.2741	-6.04	0.0001
R^2: .42; n: 115; F: 11.26			

Table 5: Relationship Between Predictor Variables and GRADE: Tenured Faculty

Variables	Parameter Estimate	t Value	PR >|T|
(Class Size ≤ 12)			
INTERCEPT	0.4066	0.91	0.3666
Pedagogy			
(Q2) Learning Experience	0.6021	3.6	0.0008
(Q7) Explanation of Assignment	-0.4905	-2.79	0.0085
Ethical & Intellectual			
(Q12) Fairness in Grading	0.5281	3	0.0048
CBK	-0.3079	-2.61	0.0131
R^2: .63; n: 40; F: 14.98			
(For Class Size > 12)			
INTERCEPT	4.05521	4.43	0.0001
Pedagogy			
(Q1) Presentation of Material	-0.4281	-2.67	0.0084
(Q8) Relevance of Assignments	0.3142	2.39	0.018
(Q6) Instructor's Expectations	-0.5305	-2.4	0.0175
Personal Characteristics			
(Q3) Stimulation of Interest	-0.4217	-3.11	0.0002
(Q10) Concern for Students	0.3739	3.74	0.0003
Ethical and Intellectual			
(Q13) Knowledge of Subject	-0.5446	-4.17	0.0001
CBK	-0.1134	-1.63	0.1054
R^2: .629; n: 149; F: 29.92			

CONCLUSION

As indicated at the beginning, this paper will not provide a definitive answer to the belief held by some that low teaching evaluations are the result of classroom rigor and that high evaluations are "bought" by giving high, unearned grades. It should be viewed as a paper to advance the discussion on this important subject. Said that, the authors will state, supported by some data, but rejected by other data, that the hypothesis that there is a relationship between high grades and teaching evaluation, due to one "buying" grades cannot be definitively rejected. The results are inconclusive and mixed. What drives high or low SPOT's is due to multiple factors. However, there appears more support in this study to reject the hypothesis that the awarding of high grades will result in high teaching evaluations than there is for accepting the hypothesis.

REFERENCES

Adams, J. (1997). Student Evaluations: The Ratings Game. *Inquiry,* 1, 10-16.

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

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

Frey, P. W. (1978). A Two Dimensional Analysis of Student Ratings of Instruction. *Research in Higher Education*, 9, 69-71.

Isley, P. and H. Singh. (2005). Do Higher Grades Lead to Favorable Evaluations? *The Journal of Economic Education*, 36, 29-41.

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

Ryan, J.M., and P.D. Harrison (1995). The Relationship Between Individual Instructional Characteristics and the Overall Assessment of Teaching Effectiveness Across Different Instructional Contexts. *Research in Higher Education*, 36, 577-94.