

September 1975

## Uncertainty, Grades and Student Evaluations

Justin D. Stolen  
*University of Nebraska*

Follow this and additional works at: <https://digitalcommons.georgiasouthern.edu/sbr>



Part of the [Business Commons](#), and the [Education Commons](#)

---

### Recommended Citation

Stolen, Justin D. (1975) "Uncertainty, Grades and Student Evaluations," *Southern Business Review*. Vol. 1: Iss. 2, Article 7.

Available at: <https://digitalcommons.georgiasouthern.edu/sbr/vol1/iss2/7>

This article is brought to you for free and open access by the Journals at Digital Commons@Georgia Southern. It has been accepted for inclusion in Southern Business Review by an authorized administrator of Digital Commons@Georgia Southern. For more information, please contact [digitalcommons@georgiasouthern.edu](mailto:digitalcommons@georgiasouthern.edu).

## Note

# UNCERTAINTY, GRADES AND STUDENT EVALUATIONS

Justin D. Stolen

There has been an increasing interest in the effectiveness of college teaching in recent years which has led to expanded efforts to evaluate that teaching. Some experts argue that this evaluation can be done best by departmental chairpersons and/or supervisory committees witnessing actual class sessions [3, p. 229]. Others argue that pre- and post-testing students' knowledge in specific classes will measure the ultimate effectiveness of the instructor [1]. While these and other suggestions have some support, another increasingly common evaluatory technique is to obtain input via questionnaires [5, p. 1069].

These questionnaires are designed to determine certain qualities of teaching and are used because students are considered capable of objectively evaluating their instructor's performance. However, recent studies analyzing student evaluations suggest that students are not particularly good judges of teaching effectiveness if that effectiveness is measured by what they have learned [6]. Such studies have raised doubts about the meaning and ultimate usefulness of student evaluations. For example, Soper [7] argues that while student evaluations measure "something," that "something" is unclear, and Muse [4] concludes that student evaluations at best indicate how well a faculty member is liked by his students.

Thus, when these evaluations are explicitly used by administrators in the evaluatory process, instructors may feel a heightened sense of uncertainty with respect to judgments of their classroom performance and may very well react to this heightened uncertainty by trying to influence student responses through higher grades. To the extent that this grade inflation becomes widespread, it can lead to competitive grade devaluation and ultimately to declining classroom standards as faculty rival each other for high evaluations. In an effort to determine whether this uncertainty on the part of the faculty is justified and whether grades do influence the evaluations, two separate questionnaires were given to 110 class sections in the College of Business Administration at the University of Nebraska at Omaha during the spring semester of 1973-1974. The results suggest that there is a high degree of uncertainty associated with interpretation of the evaluations and that average grades in

the course are one of the few variables that consistently affected both evaluations.<sup>1</sup>

Thus, this study extends beyond previous work by comparing two different evaluation instruments within a particular semester. As such, it can more adequately study both the uncertainty and grade inflation questions. Specifically, students were asked to rate certain characteristics of instructors on a scale from 1 (poor) to 5 (excellent). Within each class, single teacher evaluation measures for each of the questionnaires were then obtained by multiplying student responses for a certain question on the first instrument, or the average of certain questions on the second instrument, by twenty.<sup>2</sup> That is, for each questionnaire the measure of teacher evaluation ranged from 20 (poor) to 100 (excellent). Both measures of teacher evaluation were then stepwise regressed against a number of variables which the literature on the subject has deemed important: average grade in course; instructor's rank, age, sex, experience and educational attainment; size, time and frequency of class; and academic discipline and level of course [2]. The stepwise regression determined which variables were statistically significant for either of the two questionnaires, and then those variables were used in developing one final regression equation for each questionnaire.<sup>3</sup>

The results of these two final regressions follow:

$$E_1 = 74.60 - .43X_1 + 5.99X_2 + 3.89X_3 + .01X_4 - 7.77X_5 - 7.11X_6 + 7.17X_7$$

$$(3.64)^* (1.80)^* (2.08)^* (.01) (2.14)^* (3.42)^* (1.91)^*$$

$$-3.55X_8 + 2.78X_9 - .14X_{10} + .74X_{11}$$

$$(.82) (.40) (1.83)^* (.11)$$

and

$$E_2 = 65.23 - .38X_1 + 8.12X_2 + 5.45X_3 + 3.25X_4 - 2.04X_5 - 1.48X_6 + 3.71X_7$$

$$(4.49)^* (3.47)^* (4.13)^* (1.12) (.79) (.54) (1.41)$$

$$-6.49X_8 - .56X_9 - .08X_{10} + 1.73X_{11}$$

$$(2.52)^* (.17) (1.54) (1.08)$$

where  $E_i$  is the mean classroom teacher evaluation for the  $i$ 'th questionnaire;  $X_1$  = experience of the instructor in years;  $X_2$  = 1 if the instructor is male, 0 otherwise;  $X_3$  = average grade given in the course;  $X_4$  = 1 if the course is sophomore level, 0 otherwise;  $X_5$  = 1 if the course is junior level, 0 otherwise;  $X_6$  = 1 if the class is senior level, 0 otherwise;  $X_7$  = 1 if the course is in the law and society discipline, 0 otherwise;  $X_8$  = 1 if the

course is in the decision sciences discipline, 0 otherwise;  $X_9 = 1$  if the course is in the banking and finance discipline, 0 otherwise;  $X_{10}$  = class size;  $X_{11} = 1$  if the course is in a Monday - Wednesday - Friday sequence, 0 otherwise. The values in parentheses are *t* values, and those with asterisks indicate coefficients which are significantly different from zero. The respective multiple *R*'s for the two equations are .55 and .61 and are significant at better than the .05 level. In addition, the standard errors are 10.73 and 7.55, respectively.

Since the dependent variable in each regression is a measure of teacher evaluation, these variables should be highly correlated with each other, and the coefficients of identical independent variables in each equation should be similar in sign and magnitude. But even though there is some correlation between  $E_1$  and  $E_2$  ( $r^2 = .60$ ), it is not particularly high. In addition, even though 10 of the 11 coefficients have the same sign from one equation to another, in eight cases either only one of the coefficients is significant, or both are insignificant. Specifically, the coefficient of  $X_4$  to  $X_6$  pertaining to class level,  $X_7$  to  $X_9$  concerning discipline, and  $X_{10}$  and  $X_{11}$  relating to class size and time of day are all either not significantly different from zero for either questionnaire or significantly negative for one questionnaire and not significantly different from zero for the other. These results substantiate the hypothesis that there is uncertainty associated with the interpretation of the evaluations.

The only regression coefficients which have the same sign and are significantly different from zero in both equations are those of experience ( $X_1$ ), sex ( $X_2$ ), and grades ( $X_3$ ). The first has a significantly negative coefficient, and the latter two have significantly positive coefficients. The negative coefficient for experience is somewhat surprising. It may be that as a professor gains more experience, he loses rapport with students. That is, students may relate more to an instructor who is closer to their own level of understanding.<sup>4</sup> Maleness is positive indicating a bias against female teachers. This could be due either to business students identifying more with male teachers or to possible sampling error.<sup>5</sup>

However, the variable of particular interest for this study is grades since it is the only one of these three which can be manipulated by the instructor. For example, the results indicate that an increase of one letter grade in the average grade given in the course can increase the instructor's evaluation score by .19 points in equation one and by .27 points in equation two (on a scale of 1 to 5). The fact that the coefficient of grades is significant regardless of which questionnaire is used implies that the faculty has learned its importance. This in turn has some serious

implications regarding future grade levels and classrooms standards.

In conclusion, even though there is some similarity between the two regression equations, the statistical differences suggest that faculty are justified in questioning the ultimate meaning of the evaluations. In addition, this note suggests that regardless of what the dependant variable measures (teaching effectiveness, popularity, appearance or some other characteristic of the instructor), grade manipulation by the instructor can influence them. These results and their potential long lasting implications with regard to educational standards should be weighed heavily when consideration is made of explicit use of such evaluations in the evaluating process.

---

### REFERENCE FOOTNOTES

<sup>1</sup>While the average of the actual course grade was used instead of the average of the expected course grade, it was felt that the evaluations were given late enough in the semester so that student expectations would be accurately reflected in the actual course grade.

<sup>2</sup>On questionnaire one, the relevant question was: considering everything, how do you rate the teaching in this course? On questionnaire two, the relevant average was derived from six questions dealing with the following characteristics: analysis/synthesis, organization/clarity, instructor-group interaction, instructor-individual student interaction, dynamism/enthusiasm, and student development.

<sup>3</sup>Unless otherwise noted, all hypothesis tests in this paper were carried out at the 90% confidence level.

<sup>4</sup>This point is deserving of further study.

<sup>5</sup>Only about 5% of the instructors were female.

---

### REFERENCES

1. Anthony, William P., and Lewis, John R., "Learning Effectiveness: The Measure of Effective Teaching," *AACSB Bulletin*, 8 (October 1972), 11-17.

2. Costin, F. W. T., Greenough and Menges, "Student Ratings of College Teaching: Reliability, Validity and Usefulness," *Review of Educational Research*, 41 (December 1971), 511-535.

3. Mueller, Dorothy, "How to Evaluate Teaching," **Journal of Teacher Education**, 22, no. 2 (Summer 1971), 229-44.
  4. Muse, William V., "Focus on Learnings, Not Teaching," **AACSB Bulletin**, 10 (April 1974), 15-19.
  5. Nichols, Alan, and Soper, John C., "Economic Man in the Classroom," **Journal of Political Economy**, 80, no. 5 (September/October 1972) 1069-1073.
  6. Rodin, Miriam and Burton Rodin, "Student Evaluations of Teachers," **Journal of Economic Education**, 5, no. 1 (Fall 1973), 5-10.
  7. Soper, John C., "Soft Research on a Hard Subject: Students Evaluations Reconsidered," **Journal of Economic Education**, 5, no. 1 (Fall 1973) 22-26.
- 

Dr. Stolen is an Associate Professor of Decision Science at the University of Nebraska at Omaha.