

## ACADEMIC RIGOR AND THE STUDENT TEACHER EVALUATION PROCESS: STUDENT PERCEPTIONS

Dennis E. Clayson, University of Northern Iowa,  
Cedar Falls, IA 50614-0126; (319) 273-6015

### ABSTRACT

Two studies were conducted to look at the relationship between rigor and student evaluation of instruction. Approximately 30% of the students admitted that they had purposely inflated evaluations because an instructor gave good grades, and about 30% indicated that they had purposely lowered evaluations below what they thought the instructor deserved because tests in the course were too hard. Fifty percent had done one or the other. Students did not believe that a demand for rigor was an important characteristic of a good instructor. Further, approximately half of the students in a self-reported, published medium gave their instructor the same grade they received. The findings suggest that faculty fears that higher academic standards would lower student evaluations are not unfounded.

A recent survey conducted at a Western university found that over 65% of the surveyed faculty believed that higher standards for grades in classes would lower student evaluations (Birnbaum 2000). When asked if the student evaluation process encourages faculty to "water down" the content of their courses, 72% responded in the affirmative. Almost 49% of the faculty said they present less material in class than they used to, and about one third said they have lowered standards for students to get a passing grade (only 7% said they had raised standards).

An on-line publication recently suggested that if instructors wanted to raise student evaluations they should grade leniently (The Bucknellian Online 1998). Yet educators continue to state that rigorous academic standards do NOT significantly change student teacher evaluations (Cashin 1995; Marsh & Dunkin 1992).

Marsh and Roche (1999) referred to the idea that academic rigor would result in lower student teacher evaluations as a "presumption" that is not supported by the research. They state that the evaluation-grade expectation correlation is small ( $r = 0.2$ , approximately 4% of the variance). The authors suggest that instructors use proven methods of improving their evaluation ratings rather than, "...

the ethically dubious, counterproductive tactics (easy workloads and grading standards) that apparently do not improve teaching effectiveness of SETs (p. 518)." The debate has been complicated by research findings. Gillmore and Greenwald (1999) reported that out of six published studies that manipulated grading leniency in actual classrooms, all found higher evaluations from students in the more lenient conditions. Clayson & Haley (1990) found that in marketing classes, academic rigor was not significantly related directly to teaching evaluations. However, they found that academic rigor was significantly positively related to "learning", which was positively related to the evaluations. Rigor was negatively related to "personality" and "fairness", both of which were positively related to the final evaluation outcome. The combined overall effect of rigor on the evaluations was significant and negative. In other words, students believed that they would learn more in a course with rigor, but they also reported that rigor would decrease their evaluation of the instructor's personality and her fairness. In this study, "personality" accounted for so much of the variation of the final evaluation outcome that the overall effect of rigor was negative.

This study looks at the direct student perceptions of academic rigor and their related view of instruction.

### METHOD

Two studies were conducted. The first attempted to evaluate students' perception of rigor and its place in defining a good instructor. The second study looked at the issue of reciprocity. In other words, to what extent do students tend to evaluate the adequacy of teaching by their own grade? Do the professors' fears expressed in the survey reviewed above have a basis?

#### Method of Study One

A survey was administered to two undergraduate business classes in each of three universities. One university was in the Western states, one in the Midwest, and one in the Southwest. A total of 168 (105 female) useable questionnaires were analyzed. Each of these institutions encourages students to add written comments to the standard evaluation form.

As part of a longer survey, several questions were asked of students' perceptions of the importance of rigor in the teaching evaluation process. Rigor was defined on each questionnaire as, "... the instructor demands a lot of you." Students were asked, "Have you ever added a written suggestion [to a student teacher evaluation] that an instructor's tests were too hard?" They were also asked if they had given an average instructor, "... a higher than average teaching rating..." because the instructor gave good grades, and "... a lower than average teaching rating..." because the instructor's tests were too hard.

Further, each student was given five traits (see Table 1) that could be related to a good teacher (taken from Clayson & Haley 1990) and asked to give a percentage to each weighing that trait within the total of five. The students were also given a forced choice comparison of four traits to create a Thurstone's Paired Comparison scale.

**Results**

Thirty-five percent of the students admitted to adding written comments to evaluations stating that the instructor's tests were too hard. Twenty-nine percent said that they had given an instructor a higher evaluation than they thought the instructor deserved because that instructor gave good grades, and 29% indicated that they had given an instructor lower evaluations than the instructor deserved because tests were too hard. Forty-five percent had either added a written comment that tests were too hard and/or had purposely inflated an evaluation of an instructor that gave good grades. Forty-three percent had either added a written comment that tests were too hard and/or had purposely deflated an evaluation of an instructor that had difficult tests. Fifty percent had inflated an evaluation because of good grades they received and/or had purposely deflated an evaluation of an instructor that gave tests that were "too" hard. Table 1 shows the proportions of traits that would constitute a good instructor in the students' perception.

The Thurstone's Paired Comparison asked the students to create a scale of the "... most important characteristics of a good instructor." Learning was defined as "... how much you think you learned in the class." On a forced-choice paired comparison, learning was chosen 98% of the time over rigor, fairness was chosen 89%, and personality was chosen 75% of the time over rigor.

Thurstone's procedure creates an interval scale. These intervals were found to be (adjusting rigor to zero):

Rigor	0
Fairness	4.02
Personality	4.25
Learning	6.77

**TABLE 1**

**Proportions of Traits that Impact Good Teaching**

Trait	Av Percent	Cumulative % < 11
Knowledge	31.5	6.0
Concern for Students	21.0	19.0
Personality	18.8	34.5
Fairness	16.7	32.7
Rigor	11.5	70.8

**Discussion**

The concept of "rigor", defined as, "... the instructor demands a lot of you," was not considered by these students as an important trait to a good instructor. Almost 71% of the students reported that rigor would constitute 10% or less of the proportion of traits that would characterize a good instructor. Further, about 50% of the students indicated that they would reward or punish an instructor based on giving good grades, or giving tests that were judged to be "too" hard. The students apparently found little relationship between a knowledgeable instructor and student learning with rigor.

**Method of Study Two**

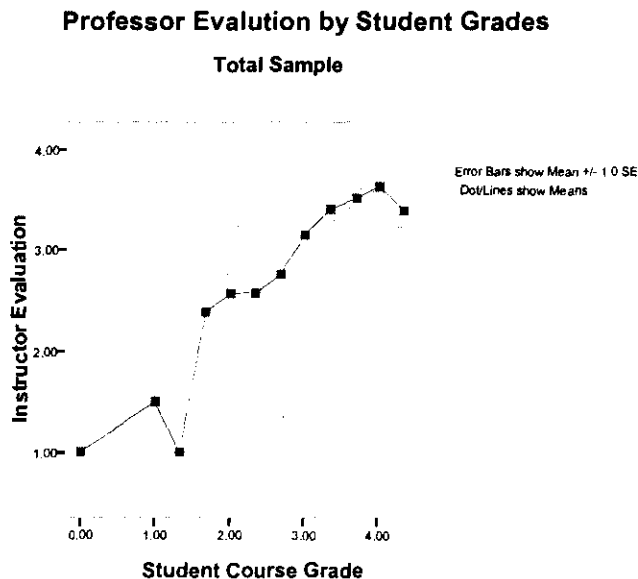
Several web sites now exist that give students an opportunity to publish an evaluation of an instructor and/or a course. One of the largest of these is Teacher Review ([www.teacherreview.com](http://www.teacherreview.com)), which currently contains evaluations of 3,011 instructors in 536 schools. Students give a letter grade to the instructor's exams, homework assignments, lectures, class projects, office hours, personality, and speaking skills. The student then gives an overall letter grade to the instructor and then can add the grade they received in the course.

Evaluations were chosen from business, social science, and physical science professors that were at least over half complete (i.e., over half of the letter grades were assigned). When more than one student evaluated an individual instructor, one evaluation was chosen at random, using a random number generator. The sample selected contained an evaluation of 809 instructors.

## Results

The relationship between the student's course grade and the student's evaluation of the instructor was analyzed. The correlation between the two was 0.472, accounting for 22.2% of the variance ( $r = 0.474$  for 182 business instructors). Of the students receiving an A in the course, 72% gave their instructor an A (87% gave either an A or B). Of the students receiving a C or lower, 67% gave their instructor a C or lower. Over 53% (53.5%) of the students gave their instructor the same grade they received ( $Z = 12.19$  that the distribution was random). The graph of the two variables is shown below.<sup>1</sup>

Figure 1



## Discussion

Students could obviously input any data they chose onto the web site. It would be expected that students who were the most computer literate and had strong feelings would be more likely to respond. There is also no external validation of the course

<sup>1</sup> Students could claim a grade of A+.

grade recorded. The average grade for the sample was 3.39 (after adjusting A+ to A). It has been this writer's experience in over ten years of measuring such things, that students have a tendency to overestimate their grades, indicating that the correlation reported here is probably underestimated (reduction of variance lowers the correlation). It is also true, however, that students during the actual evaluation process can record anything they choose. Nevertheless, there was a strong relationship between reported grades in a course and the evaluation given to the instructor.

## RESULTS AND CONCLUSION

Over a third of the students sampled reported that they had been concerned enough to add a written comment to an instructor's evaluation that the course's tests were too hard. Approximately 30% of the students admitted that they had purposely inflated evaluations because an instructor gave good grades, and about 30% indicated that they had purposely lowered evaluations below what they thought the instructor deserved because tests were judged to be too hard. Fifty percent had done one or the other.

The sampled students did not believe that a demand for rigor (as defined) was an important characteristic of a good instructor. Further, approximately half of the students in a self-reported, published medium gave their instructor the same grade they received.

The results of this study seem to support the faculty "presumptions" found in Birnbaum's survey.

## LIMITATIONS

The sample in the first study was small and not randomly chosen. The sample of the second study was self-selected and in a public medium. Generalizations of the findings are therefore questionable. The results, however, are suggestive and agree with the published literature (Greenwald & Gillmore 1997). Nevertheless in a field that literally contains thousands of published articles, these results only point to further research.

## REFERENCES

- Birnbaum, Michael H. 2000. A survey of faculty opinions concerning student evaluation of teaching. <http://psych.Fullerton.edu/mbirnbaum/faculty3.htm>.

*The Bucknellian Online* 1998. 20 ways to improve your teaching evaluations without improving your teaching. <http://www.orgs.buckness.edu/bucknellian/sp98/4-2-98/lifest/5934.html>.

Cashin, William E. 1995. Student ratings of teaching: the research revisited. *Center for Faculty Evaluation & Development*. Idea Paper No. 32.

Clayson, Dennis E., and Debra A. Haley 1990. Student evaluations in marketing: What is actually being measured? *Journal of Marketing Education* 12(Fall), 9-17.

Greenwald, Anthony G., and Gerald M. Gillmore 1997. Grading leniency is a removable contaminant of student ratings. *American Psychologist* 52(11), 1209-1217.

Gillmore, Gerald M., and Anthony G. Greenwald 1999. Using statistical adjustment to reduce biases in student ratings. *American Psychologist* 54(7), 518-519.

Marsh, Herbert W., and M. Dunkin 1992. Students' evaluation of university teaching: A multidimensional perspective. In J.C. Smart (ed.) *Higher Education: Handbook of Theory and Research*. New York: Agathon.

Marsh, Herbert W., and Lawrence A. Roche 1999. Reply upon SET research. *American Psychologist* 54(7), 517-518.