Woman C.P.A.

Volume 53 | Issue 4 Article 17

Fall 1991

Student Rating Biases: Are Faculty Fears Justified?

Robert E. Holtfreter

Student Rating Biases: Are Faculty Fears Justified?

By Robert E. Holtfreter, Ph.D., CPA

The evaluation of teacher effectiveness is a process broadly supported by students, faculty, and academic administrators in the United States. (Leventhal, Perry, Abrami, Turcotte, and Kane, 1981). The reasons for evaluating teacher effectiveness are to provide information for (1) administrators to be used for tenure, promotion, and merit decisions; (2) faculty in the form of diagnostic feedback; (3) students to be used in the selection of instructors; and (4) researchers who are studying the field of teacher effectiveness. Even though student ratings of instructors have been proven to be reliable and valid, (March, 1984; Cohen, 1981) faculty are concerned that data from student ratings are often biased or contaminated by the effects of variables outside of their control. For example, Marsh and Overall (1979) determined, in a survey of faculty, that more than half of the respondents felt that eight characteristics (from a list of 17) caused a significant bias to student ratings. The eight items included: student's grade point average (53%), class size (60%), student interest in subject before course (62%), course workload (60%), instructor popularity (63%), course difficulty (72%), reason for taking the course (55%), and grading leniency (68%). The implication of the above study is that the resultant student ratings should be adjusted for the effect of these eight characteristics or variables.

Characteristics or variables that are perceived to bias student ratings are normally referred to as extraneous variables and have been identified as having a mixed effect on student ratings. Although there are many characteristics or variables that may possibly influence student ratings, the more commonly cited ones are Class Size, Relation of the Sex of The Instructor to the Sex of The Student, Prior Interest in the Subject Matter, Administrative Leniency, Academic Field, Student Leniency, Instructor Characteristics, and Expected Grade. These extraneous variables are the focus of this article.

The purpose of this article is to review the research on these selected extraneous variables in order to determine if the fears that faculty possess concerning student rating biases are justified. The review will result in a conclusion concerning (1) the overall statistical significance of the impact that each variable has on student ratings, (2) the statistical impact that all of these variable have in aggregate on student ratings, and (3) what measures, if any, should be taken to adjust student rating data for the effects of these perceived biases.

Statistical Significance of the Impact that Individual Extraneous Variables Have on Student Ratings

The approaches authors have taken to study the statistical relationship between individual extraneous variables and student ratings have typically fallen into two major research designs. They have usually measured the impact of individual extraneous variables on either overall student ratings or on various dimensions of the student ratings instrument. Regardless of the approach taken in the research design, the results of the research are mixed. The more consistent and logical findings, though, seem to be occurring when student ratings are treated as multidimensional.

Class Size

The results of the research on Class Size is mixed. However, a large number of studies have reported statistically significant correlations between Class Size and overall student ratings ranging from negative .10 to negative .40, which, initially, indicates smaller classes are receiving the higher ratings (Caskin and Slamm, 1977a; Breadslemberg, Slindle, and Balirtu, 1977; Marsh, 1978).

In other studies curvilinear results have been reported which indicate that the typical negative relationship existing between Class Size and overall student ratings, persists to a certain class size and then reverts to a positive relationship (Marsh, Overall, and Kesler, 1979a; March, 1980b; 1983; Glass, McLaw, and Smith, 1981; Smith and Glass, 1980). Essentially, the research studies show that the smaller and larger classes are receiving the

higher student ratings relative to medium size classes which are receiving the lower ratings. There is no consensus concerning the point at which the negative relationship between these two variable ends and becomes positive. Glass reported the negative effect ending at about 40 students whereas Marsh found it to be around 200 students.

Using a multidimensional approach, Marsh (1979a; 1980b; 1983) determined that Class Size was moderately correlated with two dimensions of student ratings; "group interaction" and "individual rapport." He did not find the typical negative relationship between Class Size and (1) overall student ratings or (2) other dimensions of the evaluation instrument. Based on these results, Marsh (1984) argues that Class Size should not be treated as a bias to student ratings because its effects are legitimate and, thus, accurately reflected in the student ratings. He contends that Class Size has a moderate effect on some dimensions of effective teaching, primarily "group interaction" and individual rapport." From a logical standpoint it seems that the effect of Class Size on these two dimensions of effective teaching would normally become more negative as Class Size grows because it would become increasingly more difficult for an instructor to give the same attention to the students, both individually and on a group basis. On the other hand, relatively smaller class sizes should normally lend themselves to more effective instructor/student interaction and rapport because there would be enough time to share with all the students on an equal basis. Nevertheless, this does not explain why, at some Class Size level, student ratings have been shown to become more positive.

In summary, the research on Class Size and student ratings is very mixed and reflects a complex area of study that, in order to be understood and not misinterpreted, should obviously be viewed as multidimensional. The complexity of the problem is reflected by Cranton and Smith's (1986) research where they reported the effect of Class Size on student ratings varied tremendously, depending upon, among other things, the level of instruction (e.g. junior versus senior) and the department in which the data was collected. The multidimensional approach to research in this area has been producing the most logical results and, hopefully, future research on this extraneous variable will generate more consistent and revealing results.

Sex of the Instructor and Sex of the Student

The results of the research involving this variable are mixed, but seem to indicate that similarity of student/instructor gender has a relatively weak statistical relationship to higher student ratings. For example, Ferber and Hubert (1975) and Walker (1968) found the higher ratings were received by female instructors from female students. Walker also found that the lower ratings were received by male instructors from female students.

Prior Interest in the Subject Matter

This variable asks the question: Does a student's prior interest in a particular subject have a significant impact on student ratings? The general findings of the research in this area indicate that there is higher statistical relationship between these two variables than between any of the other extraneous variables and student ratings.

Marsh argues that Prior Interest in the Subject Matter "is a variable that influences some aspects of effective teaching (particularly Learning/Value) and these effects are accurately reflected in . . . student ratings. Higher student interest in the subject apparently creates a more favorable learning environment and facilitates effective teaching, and this effect is reflected in student ratings." As a result, Marsh argues that the relationship between Prior Interest in the Subject Matter and student ratings should be viewed as having a logical effect rather than a biasing effect.

Administrative Leniency

Administrative Leniency occurs under specific conditions and normally results in somewhat higher overall student ratings. Administrative leniency occurs when one or more of the following situations occur: (1) a faculty member tells students that the main purpose of the class evaluation is for a merit/tenure/promotion decision, (2) the student ratings are not anonymous, or (3) the faculty member is present when the evaluation is given (Feldman, 1979).

Academic Field

There have been too few studies to determine a definite trend but there is a tendency for faculty in the fine arts, languages, and humanities to receive higher evaluations than faculty in other fields such as engineering, math, and physical or social sciences (Feldman, 1978).

Student Leniency

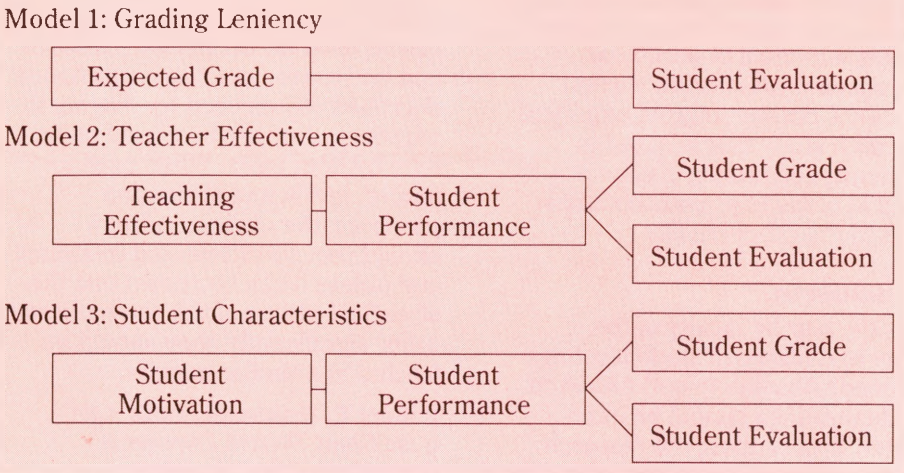
The elements of Student Leniency occur under specific conditions and normally result in noticeable but not significant differences in student ratings. The elements of Student Leniency are (1) required versus elective courses, (2) level of course, and (3) year in school. Elective courses and courses which students are taking for general interest tend to be rated somewhat higher than required courses (Marsh, 1984). Graduate level courses tend to be rated more favorably than undergraduate courses; senior courses are normally rated more favorably than junior courses; junior courses are usually rated higher than sophomore courses; and so on (Marsh, 1984; Aleamoni, 1981).

Instructor Characteristics

Instructor Characteristics have a minor impact in certain situations and have generated mixed findings with the tendency for little pattern or statistical significance. Instructor characteristics refer to Faculty Rank and Research Productivity. The relationship between student ratings and either Faculty Rank or Research Productivity appear to be positive but very weak (Aubrecht, 1979; Aleamoni, 1981). In somewhat related research, Bendig, Kulik, McKlechie (1975) and Kulik (1974) reported that communication ability was the one item that distinguished instructors with higher student ratings from those with lower ratings.

Exhibit I

Interpretive Models of Expected Effect on Student Evaluations



Expected Grade

The research on Expected Grade influence is relatively consistent in reporting a fairly significant positive correlation between class average Expected Grades and student ratings (Feldman, 1976). Brown (1976) reported that the Expected Grade effect accounted for around nine percent of the variance in student ratings. The problems associated with Expected Grade effect is not so much with the research findings, but with the interpretation of the findings.

Three different interpretive models, illustrated in exhibit 1, have emerged in the literature (Howard and Maxwell, 1980). They are the "grading leniency bias model," the "teaching effectiveness model," and the "student effectiveness model."

The "grading leniency model" suggests that faculty will receive undeserving higher student ratings if they give students undeserving higher grades, which can result in a significant upward bias to student ratings.

The "teaching effectiveness model" suggests that higher expected grades reflect higher student learning, which lends support for the validity of higher student ratings. In other words, an effective teacher normally influences higher student performance which, in turn, results in higher student grades and higher student evaluations.

The "student characteristic model" suggests that student motivation enhances student performance which, in turn, leads to higher student grades and student ratings of faculty. Both of the latter models imply that a positive correlation between Expected Grades and student ratings represent a valid outcome of student performance rather than a perceived bias in the student rating data. The research clearly supports the teacher effectiveness and the student characteristic models. Howard and Maxwell (1980) conducted two path analytic studies, the results of which strongly support the "student characteristics model" shown in Exhibit 1. By controlling for Prior Student Motivation and student progress they found a majority of the co-variation between Expected Grades and class average overall ratings was eliminated.

Using a multi-dimensional research design in his multiple regression/path analysis study, Marsh (1984) found that a major part of the relationship between Expected Grades and student ratings is spurious and largely explained by the extraneous variable "Prior Interest in the Subject Matter." Although, at first glance, this finding supports the "student characteristics model," Marsh interpreted the results as follows:

> I interpreted the results, however, as support for the validity hypothesis (teaching effectiveness model) in that prior subject interest . . . . produces more effective teaching, which leads to better student learning, better grades, and higher evaluations (student ratings).

In another multivariate study, Burton (1975) established some indirect support for the "student characteristics model" when he demonstrated that most of the eight to fifteen percent variance in student ratings was accountable by a variable referred to as "student enthusiasm." Student enthusiasm is not unlike "student motivation," which is the main characteristic of the "student characteristic model."

In summary, the research on Expected Grades and student ratings is somewhat complex but clearly indicates that the treatment of this variable as extraneous, i.e. a bias to student ratings is unfounded.

The Statistical Significance of the Correlation between Aggregating Extraneous Variable and Student Ratings

A number of studies have investigate the multivariate relationship between a wide array of perceived extraneous variables and student ratings. Initially, the research suggests that between five and nine percent of the variance in student ratings can be explained by these variables in the aggregate. But, after controlling the effect of expected grade, the variance explained in the remaining variables was negligible. For example, in their path analytical study, Stump, Freedman, and Aguanno (1979) found that only nine percent of the variance in student ratings was accountable by six perceived extraneous variables; Academic Rank, a Publication Index, Relation of the Sex of the Student to Sex of the Instructor, Class Size, Year of Experience, and Proportion of Required Courses. However, after the effect of Expected Grades was controlled for, less than four percent of the variance was accountable by these variables.

Marsh (1980) investigated the effects of sixteen perceived extraneous variable in a multiple regression/path analytic study and found

that four variables in aggregate (Prior Subject Interest, Workload Difficulty, Expected Grade, and Reason for Taking the Course) accounted for only five percent of the variance in the student ratings. The four variables were further analyzed through path analysis where it was determined that a significant portion of the relationship between Expected Grade and student ratings was nonexistent and due to other influences, thus reducing the explained variance of the remaining three variables below five percent.

Howard and Maxwell (1980) relate to this area of research best by stating "Earlier researchers seem to have jumped to a straight-forward interpretation of the relationship between grades and student ratings. As with many other complex and multidimensional psychological phenomena, careful probing can sometimes reveal obvious interpretations are misleading."

Should Student Ratings Be Adjusted?

Even though individual extraneous variables can sometimes create biases in student ratings, the research reviewed in this article strongly suggests that the statistical impact of an aggregate group of extraneous variables on student ratings is relatively insignificant. Therefore, student ratings should no be adjusted for the effects of perceived biases.

Marsh (1981) sums up this area best when he concluded that "for most of the relations (between extraneous variables and student ratings), the effects tend to be small, the results are often inconsistent, and an attribution to a bias is unwarranted." The same conclusions were drawn by Menger (1973), Centra (1979), Murray (1980), Aleannie (1981), and many more. Administrators who adjust student ratings either do not understand the literature or are playing politics by returning a favor or disfavor to particular faculty members. In the latter situation, student ratings are "justifiably" adjusted upward or downward depending upon the relative strength of the political/social relationship between the particular faculty member and the administrator.

Future Research

Although the effect of individual extraneous variables on student ratings has been well documented in other academic areas, no empirical research has been published in the accounting literature that demonstrates the effect of particular extraneous variables on student ratings. Consequently, an empirical research study that is based in academic accounting classrooms is need in order to provide comparisons to the other studies.

Conclusion

Is the fear by faculty of the perceived biases in student ratings justified? Absolutely not! Although student ratings should be interpreted with caution, the research clearly shows that perceived extraneous variables such as Expected Grades, Prior Interest in the subject, and possibly Class Size should not be viewed as potential biases in student ratings. Also, when Expected Grades are controlled for in the statistical analysis, the total actual variance explained by extraneous variables is insignificant.

Student ratings should be only part of the evaluation process and can be used for diagnostic purposes and for tenure, merit, and promotion decisions. When used for diagnostic purposes, they are best used on an item by item basis. When used for tenure, merit, and promotion purposes, the obvious outliers should receive differential treatment and unless it can be shown that the statistical means of the non-outliers differ significantly, they should be treated in a similar manner.

Robert E. Holtfreter, *Ph.D., CPA, is the Grant Thorton Professor of Accounting at Fort Hays State University in Hays, Kansas.*

Dr. Holtfreter's bibliography is too extensive to publish and is available upon request from the author.

Book Review

The Federal Data Base Finder

By Matthew Lesko
Information USA, Inc., Kensington, MD
571 pages • Price: $125 / $325 book and diskette copy

Where can the reader go to find out –

- Business patterns by geographic area and industry
- Listings of all local governments
- Trademark application files
- A directory of law enforcement agencies
- Economic and employments projections to the year 2000
- The consumer Price Index
- IRS TaxInfo

The location of all this data and much more can be found in The Federal Data Base Finder by Matthew Lesko. Lesko tells the reader everything they wanted to know about data from 14 federal departments and dozens of federal agencies, but did not know where to look.

The directory's table of contents is divided into departments, independent agencies, and branches of government. Within these divisions are contained statistics and reports compiled from the fields of trade, economics, education, science, health, and environmental issues. Each data base is listed alphabetically within each department or agency, along with an address and telephone number of who to contact, stock number (if necessary), and a description. Thousands of data bases and data files are listed. Information is often available on diskette or magnetic tape, or in searches and printouts. A handy index completes the directory, making it easy to locate specific topics.

The Federal Data Base Finder would be a valuable addition to any professional library.