

THE EFFECT OF STUDENT PERCEPTION OF INSTRUCTOR EVALUATION ON FACULTY EVALUATION SCORES

Gail Tom, California State University, Sacramento
Stokley Swanson, California State University, Sacramento
Eileen Cajucom, California State University, Sacramento
Charles Abbott, California State University, Sacramento

Student evaluations of faculty serves both as a critical criterion for administrative career decisions of retention, promotion, and tenure, and as a source of feedback to the instructor for course improvement and development. Although some studies report the validity of student evaluations as a measure of teaching performance, other studies report the effect of contextual or situation variables, such as class standing, major, and gender of the student and instructor.

Studies that have specifically investigated the effect of the intended purpose of the faculty evaluation on faculty rating report mixed findings. Some research indicates that prior knowledge of the purpose of the evaluation has no effect on faculty rating scores; other research has found that student evaluations were more favorable when students were informed of the purpose of the evaluations than when they were not. For example, certain research has indicated that students who believed that the purpose of the evaluation was administrative rated the instructor more favorably than did students who were told that the purpose of the evaluations was for course improvement.

The contradictory findings of the effect of prior knowledge of the purpose of the evaluation on faculty ratings may be due in part to the lack of control of previous studies for the initial quality or baseline performance of the instructor and, consequently, the faculty rating scores may be the result of the interaction of the baseline teaching performance of the instructor and the stated purpose of the evaluation. This study specifically tested the role of the baseline instructor performance on the effect of the purpose of the faculty evaluations on instructor ratings. It was hypothesized that differential purpose of the faculty evaluation would not affect faculty scores for instructors whose baseline performance are extreme, either very good or very poor. It was further hypothesized that the purpose of the evaluation would most significantly affect the teaching evaluations of instructors who perform at moderate levels, which is most likely the majority of the faculty at any particular school.

Three faculty from the marketing department in a large western university were selected to participate in the study. One faculty member each was selected to represent the three categories of good, moderate, and poor. Categorization of the faculty was based upon a long-term average (6 to 18 semester ratings over two to four classes) of their student evaluation ratings.

The evaluation instrument used in this study is the same as the one actually used in faculty evaluations at this school. This allows the generalization of the findings of this study to the actual faculty evaluation process used at the school.

The intended purpose of the evaluation was manipulated with one of three statements attached to the front of the survey instrument. The stated purpose of the evaluation was either to (1) assist administrators in making decisions regarding the instructor's career, or (2) serve as feedback to the instructor for instructional improvement, or (3) help students select instructors through the publications of the results.

A student administered the faculty evaluation instrument to the selected classes. Systematic random sampling was used to place students in the class in one of the three manipulated conditions. Twenty-nine students participated in each of the three classes; thus, nine or ten students were randomly placed in one of the three conditions.

The results indicate that the intended purpose of the evaluation does affect the student's evaluation of the faculty. However, the effect of the previous knowledge varies by condition. The evaluation scores for the instructor who has a history of moderate student evaluation scores was significantly higher in the administrative condition than in the instructional or public conditions. This finding supports the first hypothesis.

Contrary to the second hypothesis was the finding of significantly lower ratings for the administrative condition for the instructor who has a history of high ratings. This instructor's rating in the administrative condition was significantly lower than the ratings in either the instructional or public conditions.

The ratings received by the instructor who has a history of lower evaluation scores differed significantly in all three conditions. Students rated the instructor more favorably for administrative purposes than for instructional or public purposes. The rating of the instructor was higher for the public condition than for the instructional condition. Interestingly, prior knowledge of the purpose of the evaluation appeared to increase this instructor's ratings beyond his historical pattern. Prior knowledge did not have such a dramatic effect on the instructors with a historically high or moderate evaluation performance.

The findings of this study suggest the importance of the baseline performance of the instructor and intended purpose on student evaluation scores. The inclusion of the baseline performance variable does explain, somewhat, the diversity of previous findings. Overall, although the results are informative, they should be taken as preliminary. The generalizability of the findings would be strengthened with replications over different instructors in different disciplines and with different evaluation instruments.