Editorial

Do the Best Students Rate College Teachers the Highest?

D. A. ROBERTS, Department of Plant Pathology, University of Florida, Gainesville

"Everyone evaluates college teaching, but no one knows how to do it," wrote W. J. McKeachie (Teaching Tips, 7th ed. 1978. Heath, Lexington, MA). "Everyone" means administrators, peers, students, alumni, syndicated columnists, essayists, statesmen, politicians, and ordinary citizens, but the most of these by far are students, owing to the popularity of student evaluations among department chairmen and deans of colleges in the United States. The widely used ratings of college courses and instructors (intended to document personnel decisions, improve teaching, and assist in academic advisement) are said to be in fair agreement with evaluations made by others. Of course: members of any group are about as unlikely as those of other groups to know how to evaluate college teaching.

Never mind their questionable validity. What now of student ratings of college teaching and the effect of student achievement on those ratings? Characteristics such as age, sex, class in college, departmental major, size of class, terms in which courses are taught, and so on, seem to have no significant effects on undergraduate-student ratings of professors and courses, and the consensus is that even student grades have little or no relationship to ratings. But research results on the effects of student grades upon ratings of professors have been ambiguous. Of 26 authors whose articles were reviewed by R. I. Miller (Developing Programs for Faculty Evaluation. 1974. Jossey-Bass, San Francisco), 15 reported no correlation, six a positive one, and five a negative one between grades and ratings. K. O. Doyle (Student Evaluation of Instruction. 1975. Heath, Lexington, MA) cited 35 articles, 22 reporting grades had no influence and 13 reporting grades had a positive or negative influence. When significant relationships between grades and ratings have been found, the coefficients of correlation have been low, at least in studies of different groups of students in different courses or sections taught by different instructors, usually in a single term. My experience has been unambiguous: High positive correlations were obtained between class grade-point averages and student evaluations of my teaching of one course 16 times over 8 consecutive years.

Of 684 students in 16 classes of an upper-division college course, "Fundamentals of Plant Pathology," 524 anonymously rated me and the course on 10 items during the last week of every term. Analyses prepared later by the college administration provided, in addition to summary evaluations, the number of respondents and the percentage in each class who had given ratings in each of five scoring categories for all items: organization, clarity, concern that students learn, receptiveness, fair testing, fair grading, value of the course, gain in knowledge, overall rating of the course, and overall rating of the instructor.

Thus, it was possible to compute the average rating for each of the items, class by class. Grades and ratings were weighted to account for different numbers of students and respondents in different classes, and the weighted figures were analyzed.

Student grade-point averages in my classes were significantly correlated ($r = 0.97$) with the "overall rating of the instructor." Average student grades in each class were also significantly correlated with the respective average ratings for every one of the 10 items on the evaluation form. A second test of the relationship between grades and ratings could be made: Average ratings by students in four classes with B+ averages and in eight classes with B averages were significantly higher than those by students in four classes with C+ averages. Thus, my students in classes with the highest grade-point averages invariably rated me the highest.

I suppose superior teachers might be rated high by good and bad students alike, and extraordinarily inept teachers might receive uniformly low ratings. But for most competent college teachers of worthwhile courses, students who earn high grades might be expected to award relatively high ratings to their teachers. These students seem more perceptive than low achievers in such matters as the educational and practical significance of courses and the expositions by professors and their stimulation of student thinking. Also, high achievers seem most likely to have experienced the satisfaction of accomplishment that would make them feel kindly toward the professor and the subject. Further, the best students might just bring out the best in the instructor.

It's not clear why so many studies have led to the conclusion that little or no relationship exists between student grades and ratings of instructors, but I suspect that the relationship has been concealed by variance due to differences among courses and instructors and by the rule of anonymity. It would not be possible to study grades and ratings, student by student, without risking the identity of many respondents. My experience has been that replication, that is, evaluation of the same course and instructor at different times, revealed the strong positive correlation between student grade-point averages and their average ratings of my instruction. Replication eliminated variability due to different courses and instructors and minimized the effects of unusually high or low individual ratings upon overall variability. The number of replications needed for valid correlations is unpredictable, but the correlations between average grades and my overall average ratings were high and significant in all groups of six classes.

There has certainly been a strong positive relationship between student grades and ratings in my recent classes of "Fundamentals of Plant Pathology." Fluctuations of my ratings from 31% above to 22% below a mean derived over time probably do not reflect significant transient improvements or regressions by me. More likely, they reflect fluctuations in the caliber of students from class to class. Using student evaluations as criteria for raises, tenure, and promotion of faculty is risky business, but that risk might be somewhat reduced by appraising long-term ratings and by heeding the academic achievement of those making the ratings. As for me, give me a class of superior students every time, for I'm convinced that the best students will give me the highest ratings.