

Validity and Reliability of Student Ratings of Professors

C. E. Stufflebeam*

Introduction

Student ratings of professors and courses are relatively easy to administer and are used rather extensively in college and university classrooms throughout the United States and Canada. Because results of student ratings are frequently used in making personnel decisions, such as retention, tenure, promotion and merit evaluations, professors have a legitimate concern that the information be valid and reliable. Questions have also been raised about the effects of certain course and instructor characteristics on student ratings. The purpose of this article is to summarize the conclusions of reports from the literature regarding validity and reliability of student ratings. This investigation was conducted in conjunction with an eight year study of the effects of several student characteristics on student ratings of the instructor in an introductory course in animal science at Southwest Missouri State University. The results of that study will be reported in a subsequent article.

Evidence of research dating back to 1924 is generally positive in support of the appropriateness of using student ratings to help evaluate college instructors and instruction. In spite of this, Aleamoni (1976) listed a number of concerns voiced by faculty about their use: Can they properly measure teaching effectiveness? Are students able to make accurate judgments concerning the quality of instruction either while they are students, or after they have been away from the classroom for several years? How do teacher and student characteristics and other extraneous variables effect student ratings?

Validity

Many authorities have said that student ratings are valid if they successfully measure teaching effectiveness. According to McKeachie (1979), teaching effectiveness is the degree to which a teacher has aided in the achievement of educational goals. Logically, comparisons must be made of a number of instructors teaching the same course to similar groups of students. In this way it can be determined whether teachers whose students have learned that most were also rated highest by their students. McKeachie (1969) cited two such studies that provided persuasive evidence that student ratings were, in fact, related to teaching effectiveness in terms of student achievement in several sections of a course in chemistry.

Citing 11 studies conducted since 1969, McKeachie (1979) reported that, taken as a whole, the results confirmed earlier conclusions that teachers whose students had achieved most were also the teachers who received the highest ratings. The works of

Frey (1973) and Marsh et al. (1975) showed strong positive correlations between mean student ratings and student performance. In one study city by McKeachie, Sullivan and Skanes (1974), students were randomly assigned to instructors. This important design feature makes their report of large positive correlations between student ratings and performance of particular significance.

Further evidence from a recent study by Howard et al. (1985), shows that compared to ratings by colleagues and trained observers, ratings of teaching effectiveness by both current and former students continues to be especially effective. In fact, they found these methods to be more valid than some prior studies would suggest.

Reliability

One measure of reliability is to compare correlations between ratings at two different times by the same students. In one study by McKeachie et al. (1978), a correlation of 0.94 was found between student ratings at the end of a course and ratings by the same students 15 months later. Cited in this report were the results of a 1951 study in which a correlation of 0.6 was found between ratings made by current students and those of alumni who had graduated 10 years earlier. Overall and Marsh (1980) reported a correlation of 0.59 between ratings of current students and ratings of the same students after being out of college for at least one year. The latter study involved 1374 students in 100 classes.

The results of a study by Firth (1979) involving 723 students seem to support the conclusion that student ratings are quite stable over time and suggest that the added perspective of at least one year out of college does not alter the ratings given at the end of the course. The results of a review by Aleamoni (1981) show that alumni who had been out of school for five to 10 years rated instructors much as did students who were currently enrolled. In an earlier review, Costin and his associates (1971) had concluded that students can rate classroom instruction with a reasonable degree of reliability. Their evidence argues against the idea that student opinion is difficult to analyze or that they might be influenced by particularly good or bad experiences in the classroom.

Other measures of reliability include the degree to which student ratings of a teacher agree from semester to semester, and from course to course. According to results of research cited by McKeachie (1978), different students taking the same course taught by the same teacher, tend to rate the instructor in much the same way. Teachers' ratings also tend to be similar from course to course. Students even appeared to rate instructors similarly when different but equivalent rating forms were used.

Stufflebeam is professor of Animal Science, Southwest Missouri State University, Springfield, MO 65804.

Reliability of student ratings of instructors is apparently higher than that of colleagues. Central (1975) compared ratings based on classroom visits by peers, to student ratings. Three different peers evaluated several teachers on two occasions with very little agreement among their ratings. The correlations between their ratings and those of the students were also low. This evidence argues against the idea that peers are better able to measure teaching effectiveness than are students. Doyle and Crichton (1978) found the reliability of student ratings to be substantially higher than that of colleagues. Braskamp (1980) listed a number of problems associated with colleague ratings, especially classroom visits. He concluded that classroom visits by peers might be useful for self-improvement, but probably not very beneficial for promotion and tenure purposes.

Course and Teacher Characteristics

A number of the concerns about validity and reliability of student ratings of instructors voiced by professors relate to various characteristics of the course and instructor that may unduly influence the outcome of student ratings. A considerable amount of data has been collected regarding the influence of such variables.

Sex of the instructor is one variable that apparently has very little effect on ratings made by students according to a review of literature made by McKeachie (1979). He found, however, that certain personality traits did seem to influence student ratings. For example, teachers scoring high in personality traits such as extraversion, intuitiveness and "feeling" received higher student ratings. Highly rated teachers also seemed to be more dynamic, amicable and intellectual. Erdle *et al.* (1985) reported that about one-half the relation between personality and teaching effectiveness was mediated by classroom behavior. He suggested that instructor personality is reflected in specific classroom teaching behaviors which in turn are validly rated by students. Abrami and Mizener (1985) studied the possible effect of perceived similarities in attitudes between students and instructors on a variety of issues. They did not find this to be a substantial source of bias in ratings.

With respect to effects of class size on student ratings, some studies have shown differences, some have not. Aleamoni and Graham (1974) found no significant differences in ratings by students in three different categories of class size: less than 20, 20 to 40, and more than 40 students per class. However, Elmore and Pohlmann (1978) found that smaller classes resulted in higher ratings. In a review of about 30 studies by Feldman (1978), approximately one-third showed no relationship between class size and ratings, while about two-thirds reported small negative correlations (the smaller the class size the higher the ratings). More recently, Feldman (1984) again reported a small negative relationship between class size and ratings. Among the several reports, there were in-

consistencies as to the definition of small, medium or large with regard to class size.

Students enrolled in higher level courses had a tendency to rate their instructors higher (Aleamoni, 1980). In 23 studies reviewed by Feldman (1978), small positive correlations between class level and student ratings were reported. However, 11 other studies showed no differences. In the same review, seven studies showed no differences between the time of day classes met and student ratings. While some differences were reported in four other studies, results did not fit a consistent pattern. Student ratings also tended to be lower in required courses than in elective courses according to Centra (1978) and Aleamoni (1980).

A number of sources, including Braskamp *et al.* (1984), have presented evidence that students are reliable sources of information for describing student workload, student-teacher relationships, professional and ethical behavior of the instructor, what they learned, and how the teacher communicated. However, they do not seem to be in the best position to judge relevance and recency of subject matter, or knowledge and scholarship of professors. The latter factors can better be judged by colleagues who can also evaluate course content, assignments, testing and grading practices, text selection, and student achievement. This suggests that in the complete evaluation of instructors, information should be collected not only from students, but from colleagues, records, visuals, class outlines, administrators, the instructor, and other relevant sources, (Braskamp *et al.*, 1983).

Conclusions

In general, the literature is highly supportive of the significance of student ratings as accurate reflectors of student attitudes. Evidence indicates that as a group, students are very perceptive and reliable judges of the instructional setting. They tend to be discriminating and consistent in their judgements of instructors and instruction. Properly conducted ratings using well-designed instruments are both valid and reliable.

Literature Cited

- Abrami, P. C. and D. A. Mizener. 1985 "Student/Instructor Attitude Similarity, Student Ratings, and Course Preference." *Journal of Educational Psychology*. 77:693.
- Aleamoni, L. M. 1976. "Typical Faculty Concerns about Student Evaluations of Instruction." *NACTA Journal*, 20(1):16.
- Aleamoni, L. M. 1980. "The Use of Student Evaluations in the Improvement of Instruction." *NACTA Journal*, 24(3):18.
- Aleamoni, L. M. 1981. "Student Ratings of Instruction," from J. Millman (ed). *Handbook of Teacher Evaluation*. Sage Publications, Beverly Hills.
- Aleamoni, L. M. and M. H. Graham. 1974. "The Relationship Between CEQ Ratings and Instructors' Rank, Class Size, and Course Level." *Journal of Educational Measurements*. 11:189.
- Braskamp, L.A. 1980. "What Function Can Colleagues Have in the Evaluation of Instruction?" *NACTA Journal*, 24 (2):16.
- Braskamp, L.A., D.C. Brandenburg, E. Kohen, J. C. Dry and P. W. Mayberry. 1983 "Guidebook for Evaluating Teaching: Part 1: Rationale and Principles." *NACTA Journal*, 28 (1):19, 28 (2):27 and 28 (4):27.
- Centra, J.A. 1975. "Colleagues as Raters of Classroom Instruction." *Journal of Higher Education*, 46:327.

Centra, J.A. 1978. "Using Students Assessments to Improve Performande and Vitality." *New Directions for Institutional Research*. 20:31. Cited by McKeachie, 1979.

Costin, F. W., W. T. Greenough and R. J. Menges. 1971. "Student Ratings of College: Reliability, Validity and Usefulness." *Review of Educational Research*. 41:511.

Doyle, K. O., Jr. and L. I. Crighton. 1978. "Students, Peer and Self Evaluations." *Journal of Educational Psychology*. 70:815.

Emore, P. B. and J. T. Pohlman. 1978. "Effect of Teacher, Student and Class Characteristics on the Evaluation of College Instructors." *Journal of Educational Psychology*. 70:187.

Erdle, S., H. G. Murray and J. P. Rushton. 1985. "Personality, Classroom Behavior and Student Ratings of College Teaching Effectiveness: A Path Analysis." *Journal of Educational Psychology*. 77:394.

Feldman, K.A. 1978. "Course Characteristics and College Students' Ratings of Their Teachers: What We Know and What We Don't." *Research in Higher Education*. 9:199.

Firth, M. 1979. "Impact of Work Experience on the Validity of Student Evaluations of Teaching Effectiveness." *Journal of Educational Psychology*. 71:726.

Frey, P. W. 1973. "Student Ratings of Teaching: Validity of Several Rating Factors." *Science*. 182:83.

Howard, G.S., C.G. Conway and S.E. Maxwell, 1985. "Construct Validity of Measures of College Teaching Effectiveness." *Journal of Educational Psychology*. 77:187.

Marsh, H. W. 1984. "Student Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases, and Utility." *Journal of Educational Psychology*. 76:707.

Marsh, H. W., H. Fleiner and C. S. Thomas. 1975. "Validity and Usefulness of Student Evaluations of Instructional Quality." *Journal of Educational Psychology*. 67:833.

McKeachie, W. J. 1969. "Student Ratings of Faculty." *AAUP Bulletin*. 55:439.

McKeachie, W. J. 1979. "Student Ratings of Faculty: A Reprise." *Academe*. 65:384.

McKeachie, W. J., Yi-Guang Lin and C. N. Mendleson. 1978. "A Small Study Assessing Teacher Effectiveness." *Contemporary Educational Psychology*. 3:352.

Overall, J. U. and H. W. Marsh. 1980. "Students' Evaluations of Instructors: A Longitudinal Study of Their Stability." *Journal of Educational Psychology*. 72:321.

Sullivan, A. M. and G. R. Skanes. 1974. "Validity of Student Evaluation of Teaching and the Characteristics of Successful Instructors." *Journal of Educational Psychology*. 66:584.

Non Print Media Reviews

contact:

Gary D. Lemme, Chair
Media Review Board,
Plant Science Dept.
South Dakota State U.
Brookings, SD 57007
(605) 688-4586

1988

NACTA* Judging Contests and NACTA Judging Conference

will be held at the
University of Wisconsin-River
Falls Campus
River Falls, Wisconsin
April 29 and 30, 1988
Team Coaches Meeting
April 28, 1988

Contests to be held are:

Horses	Larry Kasten (715) 425-3704
Dairy	Perry Clark (715) 425-3704
Livestock	Dewey Wachholz (715) 425-3150
Mechanics	Jerry Nechville (715) 425-3728
Soils	Larry Meyers (715) 425-3395
Crops	Steve Carlson (715) 425-3989
Horticulture	Lanny Neel (715) 425-385

For information and Entry Forms:
Contact:

College of Agriculture
210 Ag Science Building
University of Wisconsin-River Falls
River Falls, WI 54022
715-425-3784
JOIN NACTA!

* National Association of Colleges and Teachers of Agriculture

34th Annual NACTA Conference
June 26-29, 1987
Oregon State University

The NACTA International Programs Committee is developing sets of 2x2 slides on international agriculture to make available to interested NACTA members for classroom instruction purposes. The format for these slide sets will be geographical and by disciplines (e.g. Animal Agriculture of South America). Any NACTA member or friend wishing to share high quality slides with the committee, please identify each slide before mailing to the chairman of this committee at the address below. We are hoping that this slide sets will aid our NACTA members to increase awareness of international agriculture in the college classroom.

Send to: Robert A. Godke, Chrm.
Animal Science Dept.
Louisiana State University
Baton Rouge, Louisiana 70803