

## STUDENT EVALUATIONS OF TEACHING

SUSAN J. FRANCES, *Department of Psychology*

The procedure in which students evaluate their instructors on anonymous questionnaires is a familiar one to faculty at H.S.U. It has been mandatory for most junior faculty for years, and a recent C.S.U.C. systemwide decision has required even full professors to collect yearly evaluations. Although we all participate in this process, there does not seem to be much faculty consensus about its significance. Issues regarding interpretation and utilization of student evaluation data remain controversial and largely unresolved.

In an attempt to clarify my own thinking about student evaluations, I decided to read some of the educational psychology research literature on the topic. This is an extensive literature in which many of the controversies surrounding student ratings are addressed. Not wanting to rely completely on other people's findings, I also worked with my Psychology Department colleagues Mary Gruber and Marilyn Brown on a statistical analysis of all the student evaluation forms collected in Psychology classes during 1979.

Through these endeavors I did arrive at some tentative conclusions about the merits and shortcomings of student evaluations of instruction. The purpose of this article is to share my current thinking, solicit other perspectives, and encourage further faculty discussion regarding the entire student evaluation process. I will not attempt to review all the relevant published research or to report all the findings of the Psychology Department study. A selected bibliography is included at the end of the article, and I would be happy to provide specific references and/or the Psychology Department report upon request.

The most pertinent evidence regarding the validity of student evaluations has come from research in which student ratings of instructors are correlated with some valued educational outcome

among those instructors' students. The most common research design involves multi-section courses in which common educational objectives and examinations are specified for all sections. Different instructors are assigned to different sections. At the end of the term each instructor's average student rating score is correlated with the average final exam score (or average score on a measure of student interest in subject matter) for his/her section.

Although some of the studies designed along these general lines can be faulted on methodological grounds, others have been rather neatly done. Moreover, the majority of these studies have yielded such consistent findings, that, to me, the results are convincing. **Again and again, a significant positive correlation has been found between instructors' student ratings and the achievement of their students on final exams.** Fewer studies have examined affective outcomes such as increased interest in subject, and the findings have varied somewhat depending on the measures employed, but here too the general trend has been to report positive associations with student ratings.

The circumstance of multi-section courses with fixed educational objectives and examinations (and possibly fixed reading lists and weekly assignments too) is not a common one at H.S.U., however. Typically faculty here are responsible for planning a course, as well as teaching particular sections of it. Are students as good at evaluating the content of a course as they are at evaluating the implementation of a carefully designed course plan?

Another large body of research speaks to this issue. In studies following the "Dr. Fox paradigm," students are shown videotaped lectures and then asked to evaluate the lecturers. The original Dr. Fox was an actor who presented

various versions of a lecture. The versions were systematically varied in terms of both content value and expressiveness of delivery. **The general finding was that Dr. Fox received higher student ratings for expressively delivered lectures with minimal content than for lectures with high-quality content delivered in a monotone.** Similar findings have been obtained from subsequent replications and with subject populations other than college students. Even professors and professionals tend to rate expressive, minimal-content lectures more highly than non-expressive, high-content lectures *as long as the lecture is not in their field of expertise.*

These two bodies of research suggest to me that student ratings provide a meaningful indicator of instructor performance when course content and lecture content is appropriate. Student ratings do *not* provide a means for assessing the quality of course content or for assessing the instructor's competence in an academic area. This makes sense. Why should students taking a course for the first time be able to assess the validity, currency, or relevance of the material presented or to judge if "enough" material was covered? On the other hand, students can certainly tell very well whether whatever the instructor presented was presented in a manner which kept them awake and interested. They can also report on other aspects of their experience in the class such as the clarity of the assignments or the openness of the instructor to questions, which probably contribute to desirable educational outcomes.

In practice, student ratings of instructors correlate reasonably well with faculty ratings of the teaching ability of the same instructors, according to several studies. Other research indicates that student ratings also correlate quite well with administrative ratings and with alumni ratings of faculty teaching ability.

Many different sorts of items have been used on student evaluation questionnaires. Ratings on global evaluation items, such as "all around teaching ability" correlate most highly with student achievement measures in the multi-instructor, multi-section course research. Factor analytic studies of evaluation items generally find that global items cluster with other items, such as "interesting presentation of material" and "clarity," on a dimension which is sometimes named "instructor skill." Other items often cluster on factors named "rapport," "course structure," and "course difficulty."

In our factor analysis of items on 1292 Psychology faculty evaluation questionnaires, all seven items which asked students to rate their instructors loaded on a single factor. Other items which asked students to rate the amount of work required, their level of class participation, their interest in the subject, etc., did not load on this factor. The item with the highest loading on the instructor evaluation dimension was "overall teaching effectiveness." A close second was "clarity of presentation." This suggests that students do not rate an instructor highly on the global evaluation items primarily because he/she is friendly or the course is easy.

There are some factors extraneous to teaching skill which may affect global evaluation, however. Class size and academic field are two factors which have been identified in many studies. Teachers of smaller classes tend to receive higher ratings, most notably when the class consists of twelve or fewer students. Teachers in the humanities also tend to receive higher ratings than teachers in social or physical sciences.

Some research also suggests that an instructor's grading practice may have a small, but statistically significant, effect on his/her overall student ratings. Students who expect higher grades tend to give higher ratings, and instructors of classes in which many high grades are expected tend to receive slightly higher average ratings than other instructors. A positive correlation between expected grade and instructor rating was found in the Psychology Department evaluations, though, as in other studies, its magnitude was small.

All studies of the distribution of student ratings have found them to be markedly skewed. Most studies report that over half the faculty have mean ratings of four or more on typical five-point rating scales for overall teaching effectiveness, where five is the highest score. Averaging scores across the 1292 Psychology Department forms, we found that the mean rating on every instructor evaluation item was above four, where five was "outstanding" and one was "unsatisfactory." The most straightforward interpretation of these results is a cheerful one: apparently most college students are pleased with their instructors.

Information gained through the student evaluation process is commonly used in two rather different ways. Student evaluations are supposed to provide feedback which will assist instructors in improving their teaching performance. They are also supposed to provide relevant data to those

charged with evaluating the instructor for purposes of reappointment, tenure, or promotion. Since there is no reason to assume that the same information will be equally useful in both areas, I will discuss them separately.

With respect to their feedback function, most research has been found that student evaluations have a positive impact on subsequent instructor performance (as measured by subsequent student evaluations) under only two conditions. The first is the condition in which instructors have an especially unrealistic view of their students' opinions of them. According to several studies, when instructors' self-evaluations of their teaching are considerably better than student ratings, exposure to the student ratings leads to change in instructional practices.

The second condition under which student feedback has been found to produce instructional improvement is one in which the feedback is supplemented by consultations from colleagues. Some researchers have arranged for instructors to discuss their student evaluations with more experienced colleagues. They have found that after a consultant provides suggestions and encouragement, subsequent student ratings tend to be more positive. In the absence of consultation, if instructors have realistic perceptions of student opinion, the effect of student feedback on subsequent instructor performance is negligible.

The research findings match my own observations. Receiving student ratings in the "good" to "outstanding" range may provide a morale boost to many faculty members, but it does not provide much guidance related to instructional improvement. Receiving positive student ratings in some format which makes it clear that "good" is only "average" among one's colleagues (e.g., departmental print-outs of mean ratings) may undercut the morale boosting value of the feedback, but it does not provide any suggestion for constructive change. Receiving ratings in the "poor" to "average" range, if one is already aware of student opinion, may contribute to further demoralization, but it is still of little help in improving teaching performance. **In sum, I suspect that the student evaluation process, as it usually is structured, is not particularly helpful to most faculty.**

The other major function of student evaluations is to provide information relevant to teaching effectiveness to colleagues of faculty personnel committees and to administrators. According to a 1979 article in *The Chronicle of Higher Education*, the

consideration of student evaluations in faculty personnel decisions is now widespread among American colleges and universities. Student evaluations affect tenure and promotion decisions even in research-oriented institutions, such as U.C. Berkeley. Considerable faculty controversy has arisen regarding this practice.

My opinion, which is probably as controversial as anyone else's, is that consideration of student evaluation questionnaires in personnel deliberations is legitimate, as long as major errors of interpretation are avoided. In the first place, it must be understood that positive student evaluations do not, by themselves, demonstrate teaching effectiveness. Colleague evaluations of instructor expertise and of course syllabi, assignments, lectures, etc., provide information essential to the interpretation of student ratings. It should also be understood that small differences among instructors in average student rating scores are unlikely to be meaningful. Since a variety of extraneous factors, such as class size, academic field, and grading practices, can affect ratings, it is nonsensical to play "numbers games" by ranking instructors on the basis of average scores.

When these errors are avoided, student evaluations provide information highly relevant to the assessment of teaching effectiveness. Both the research evidence and common sense suggest that positive student response to an instructor tends to be associated with desirable educational outcomes. Student evaluation data must be carefully interpreted, but so must the other types of information included in personnel files. No one argues that academic publications should be omitted from consideration in personnel reviews merely because average number of pages published per year leaves something to be desired as a measure of scholarly achievement.

I believe that all aspects of the student evaluation process could be most fruitfully handled within the context of particular disciplines and teaching areas. Questionnaire design, administration, and interpretation could be tailored to suit specific teaching situations. This would improve the quality of the data considered in personnel reviews.

By paying more attention to the student evaluation process and taking a more active role in structuring it, faculty might also be able to minimize the anxiety engendered by the process and to maximize its feedback value. I can imagine various arrangements which would allow me to sit down with colleagues at the end of each quarter and comfort-

ably discuss how my classes had gone, what the student evaluations suggested, and ideas for instructional improvement. Some research suggests that professors' student ratings tend to begin a slow downhill slide after about twelve years of teaching experience. Perhaps a regular consultative arrangement could serve to counter this trend by encouraging both experienced and inexperienced faculty to reflect upon teaching methods and experiment with fresh approaches.

It is my conviction that any changes in the student evaluation process or in utilization of student feedback should be initiated by the faculty directly affected by those changes. They know best what is likely to prove useful. It is also my conviction that faculty should take some initiative in this area, rather than merely assuming a defensive posture regarding student demands for greater influence in personnel decisions or administrative moves toward post-tenure review. Less wasted effort and more productive outcomes are likely to result if faculty exercise leadership in matters related to instructional quality. I think that the student evaluation process could be designed to have a more positive effect on the practice of teaching.

#### SELECTED BIBLIOGRAPHY

Aubrecht, J.D. Are student ratings of teacher effectiveness valid? Manhattan, KS: Center for Faculty Evaluation and Development, Kansas State University, 1979.

Centra, J.A. Using student assessments to improve performance and vitality. In W. R. Kirschling (Ed.), *New Directions for Institutional Research: Evaluating Faculty Performance and Vitality*, no. 20. San Francisco: Jossey-Bass, 1978.

Marsh, H. W. and Overall, J. U. Validity of students' evaluations of teaching effectiveness: cognitive and affective criteria. *Journal of Educational Psychology*, 1980, 72, 468-475.

McKeachie, W. J. Student ratings of faculty: a reprise. *Academic Bulletin of the A.A.U.P.*, 1979, 65, 384-397.

McKeachie, W. J., Lin, Y., Daugherty, M., Moffett, M.M., Neigler, C., Nork, J., Walz, M., and Baldwin, R. Using student ratings and consultation to improve instruction. *British Journal of Educational Psychology*, 1980, 50, 168-174.

Rotem, A., and Glasman, N.S. On the effectiveness of students' evaluative feedback to university instructors. *Review of Educational Research*, 1979, 49, 497-511.

Schultz, C. B. Some limits to the validity and usefulness of student ratings of teachers: an argument for caution. *Educational Research Quarterly*, 1978, 3, 12-27.

Student ratings of professors exert growing influence. *The Chronicle of Higher Education*, Oct. 29, 1979, 12.

Stumpf, S. A., and Freedman, R. D. Expected grade covariation with student ratings of instruction: individual versus class effects. *Journal of Educational Psychology*, 1979, 71, 293-302.