

November 25, 1996

There are probably more studies of student ratings than of all of the other data used to evaluate college teaching combined. Although one can find individual studies that support almost any conclusion, for a number of variables there are enough studies to discern trends. In general, student ratings tend to be statistically reliable, valid, and relatively free from bias, probably more so than any other data used for faculty evaluation. Nevertheless, student ratings are only one source of data about teaching and must be used in combination with multiple sources of data if one wishes to make a judgment of all the components of college teaching. Further, student ratings are data that must be interpreted. We should not confuse a source of data with the evaluators who use the data to make a judgment.

(Cashin, 1988)

Principles for the use of student ratings for faculty evaluation and development

In using and interpreting student ratings, committees, department chairs and other administrators should be guided by these principles:

1. Student ratings are a valuable source of data about effective teaching. (Research shows that they are highly reliable and stable, and are relatively valid against a variety of indicators).
2. Student ratings must be supplemented with other sources of information about teaching, such as: syllabi; tests; grade distributions; faculty self-evaluations; and reports of class observations by colleagues.
3. Student ratings are only one tool for helping to improve teaching. Ratings can be a valuable source of information that aids in instructional improvement. They may help to identify particular aspects of a course that merit attention (e.g., test and exams, organization, student interaction, etc.). However, while serving this "diagnostic" function, standard rating forms are not very "prescriptive" in terms of specific suggestions to the instructor for making adjustments. In this regard, the role of one's peers or chairperson, confidential interviews with students, responses to open-ended questions, and teaching workshops are probably more valuable. Faculty members should be encouraged to do their own confidential surveys during the course of the semester.
4. Administrators and peer committees who rely on student ratings have a responsibility to be educated in the interpretation of these data. Extensive research has been done on student ratings and those who use these data should be aware of the strengths and limitations of student ratings. Among other things, the research suggests the following principles:

Standardized procedures should be carefully developed and scrupulously followed for the administration of student rating forms. For example, the forms should be anonymous, the instructor should not be in the room when the forms are administered; the forms should be distributed toward the end of the semester; and students should not be unduly rushed when completing them.

When using student ratings for evaluative purposes, look primarily at data that include multiple classes or sections across several semesters, and from classes that have at least 10 raters and two-thirds of the class present. Data on a few classes or a few students are likely to be unreliable.

Do not overanalyze the data. Small differences between one instructor or another or between one section and another may be meaningless.

Look at the results in terms of relevant comparisons; the absolute numbers in student ratings are of only limited value. Department chairs should furnish faculty members with reports that allow faculty members to compare their own results to that of the department as a whole.

Take into account the fact that some courses typically receive higher or lower evaluations, and do not inappropriately reward or penalize instructors for these differences. For example, larger classes typically receive slightly lower ratings, as do required courses and courses outside of a student's major area. Upper-level courses (especially graduate level courses) tend to receive higher ratings.

When using student ratings for evaluative purposes, focus primarily on "global" evaluative items about overall instructor performance and overall course satisfaction. While these questions are appropriate for evaluative purposes, they are of much less value for