

Overview of Student Ratings: Value and Limitations

The IDEA Center
www.idea.ksu.edu



Student ratings of instruction have been the subject of over 2,000 published research studies. The vast majority of these offer assurance that ratings possess acceptable reliability (if there are at least 10 raters). Ratings have also been shown to be reasonably valid in a number of different ways. Perhaps the most convincing of these are the studies of multiple sections of large courses in which student ratings were highest for instructors whose students made the best grades on a common final examination.

But there are no perfectly valid measures in education, including student ratings of teaching effectiveness. For one thing, no rating form can include all possible course objectives or teaching methods; ratings of courses with highly unique purposes and approaches may ignore salient variables. It is also true that ratings are colored by the attitudes students bring to the class. In some classes, nearly all students are highly motivated and eager to learn (especially in graduate or professional classes, but also in some undergraduate classes); in such cases, student ratings of learning outcomes are almost always favorable even if the instructor's teaching skills are marginal. Similarly, if most students take the class only because it meets a requirement, or because it was offered at a favorable time, or for some other non-academic reason, both learning outcomes and ratings of instructors can be negatively affected. The IDEA system attempts to "level the playing field" by taking such extraneous factors into account (through "adjusted" ratings). But there are probably additional factors of this type which have not yet been studied and which, therefore, have not been considered in your report.

In addition to extraneous influences on student ratings, two other limitations need to be acknowledged. First, student ratings are subject to the same shortcomings that plague all rating processes. Second, there are a number of important facets of teaching excellence which students are simply unqualified to judge.

Limitations of Ratings. Of the several weaknesses inherent in the rating process, two merit special attention. The first is the "halo effect," so-called because it describes the tendency of raters to form a general opinion of the person being rated and then let that opinion color all specific ratings. If the general impression is favorable, the "halo effect" is positive and the individual receives higher ratings on many items than a more objective evaluation would justify. The "halo effect" can also be negative; an unfavorable general impression will lead to low marks "across the board", even in areas where performance is strong. Because of this effect, student ratings make less differentiation between "strengths" and "weaknesses" than is desirable.

A second weakness is the "Error of Central Tendency." Most people have a tendency to avoid the extremes (very high and very low) in making ratings. As a result, ratings tend to pile up more toward the middle of the rating scale than might be justified. In many cases, ratings which are "somewhat below average" or "somewhat above average" may represent subdued estimates of an individual's status because of the "Error of Central Tendency."

Limitations of Students as Raters. Although, when appropriately adjusted and averaged, students ratings of their own learning and of the instructor's techniques have acceptable validity, students are not qualified to judge many other factors which characterize excellent instruction. They can't judge, for example, the appropriateness of the instructor's objectives, the relevance of assignments or readings, the degree to which subject matter content was balanced and up-to-date, or the degree to which grading standards were unduly lax or severe. These, and other dimensions of teaching excellence, are important to a comprehensive evaluation of instructional effectiveness; but methods other than "student ratings" are needed to assess them.

Student ratings can be valuable indicators of teaching effectiveness, and they can help guide improvement efforts. But they are most useful when they are a part of a more comprehensive program which includes additional evaluation tools and a systematic program for faculty development.