

# Suggestions for using student ratings of instruction in the evaluation of teaching

Prepared by J. David Perry, BEST/ISS, for the Office of Academic Affairs & Dean of the Faculties

---

Student ratings of instruction (commonly referred to as course evaluations) are useful for course improvement, and are also widely used in tenure, promotion, and salary decisions. The following suggestions are drawn from research and best practices in the administrative use of student ratings.

## General guidelines for using student ratings for administrative purposes

---

- **Use multiple sources of data.** Student ratings should not be the only source of data on which teaching is judged. Other sources such as colleague observation reports, reviews of course materials, or sample student products should be drawn on to provide a balanced view of teaching.
- **Include several courses.** One author recommends at least five sets of ratings, each based on at least fifteen students. Another suggests including a minimum of two courses each from three different terms.
- **Consider the number and proportion of students responding.** Classes with fewer than ten students or in which fewer than two-thirds of the students completed the evaluation may not constitute a reliable sample.
- **Rely mainly on global items.** For personnel decisions, most authors suggest relying on a few global or summary items, such as overall ratings of the course and instructor. Such items have been shown to correlate more highly with student learning than more specific items.
- **Don't over-interpret the data.** Rating forms typically use a five-point scale. Due to measurement error, item averages that differ within +/- .3 may not really be different. When making comparisons among faculty, most authors suggest sorting student ratings data into several groups, rather than considering exact scores.
- **Don't give undue weight to open-ended comments.** Student written comments can be helpful for course improvement; however, some authors have expressed concern about their administrative use, mainly because of the difficulty in determining the representativeness of written comments.

## Course and student characteristics that may influence ratings

---

- **Disciplinary differences.** Classes in mathematics and the natural sciences tend to receive lower ratings than those in other disciplines. It is not clear whether this is due to the difficulty of the content or differences in teaching strategies.
- **Student motivation.** Students who are required to take a course tend to rate it lower than those who elect to take it. Some studies suggest that majors give higher ratings than non-majors, but this finding is not consistent.

- **Course level.** Higher-level courses tend to get higher ratings than lower-level courses. Some studies have found a similar relationship with student standing; i.e., graduate students give higher ratings than seniors, seniors higher than juniors, and so on.
- **Class size.** Some studies report that smaller class sizes tend to give higher ratings than large classes; others have found a U-shaped relationship in which the highest ratings were obtained in classes under 35, followed by those over 100, with the lowest ratings for classes of 35-100. Most authors have concluded that, if a relationship exists between class size and ratings, it is too weak to merit much consideration.

### **Some specifics about the Multiple-Option Course Evaluation System (Multi-Op)**

- **General.** While there is no standard student ratings system at IUB, Multi-Op is the most widely used system. Multi-Op provides the ability for departments or individual instructors to create custom forms from a catalog of 183 items, plus up to fifteen instructor-written items.
- **Multi-Op averages.** Multi-Op means for Likert-scale items are based on a scale of 0 (Strongly Disagree) to 4 (Strongly Agree). Note that, while the scale midpoint is 2.0, the campus mean for most items is in the range of 2.8 to 3.3. As noted above, small differences in item averages may not be meaningful.
- **Reference groups and percentile ranks.** In addition to item means, the Multi-Op report provides percentile ranks in comparison to the campus and a designated reference group. The reference groups correspond to the professional schools and, in the College, to one of four disciplinary groups: Physical Sciences and Mathematics, Humanities, Social Sciences, and Natural Sciences.
- **Relationship between means and percentiles.** Because item means tend to cluster around 3.0, changes in means around that value will result in larger changes in percentile rank than changes further from the mean. For example, the difference between means of 3.1 and 3.3 might be 15 percentile rank points, while the difference between means of 2.1 and 2.3 might be only 5 percentile points.

### **Sources:**

- Arreola, R. A. (2000). *Developing a comprehensive faculty evaluation system (2<sup>nd</sup> edition)*. Bolton, MA: Anker Publishing Company, Inc.
- Braskamp, L. A. & Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco: Jossey-Bass.
- Cashin, W. E. (1999). Student ratings of teaching: Uses and misuses. In P. Seldin (Ed.), *Changing practices in evaluating teaching*, 25-44. Bolton, MA: Anker Publishing Company, Inc.
- Theall, M. & Franklin, J. (1991). Using student ratings for teaching improvement. In M. Theall & J. Franklin (Eds.), *Effective practices for improving teaching. (New Directions for Teaching and Learning, No. 48, Winter 1991)*, 83-98. San Francisco: Jossey-Bass.