For Faculty and Administrators: Guidelines for Interpreting Student Ratings

Student course ratings are widely accepted as a means for gathering valuable information for both evaluating and improving teaching. Extensive research demonstrates their reliability and validity. However, information from student ratings is only reliable and valid when interpreted properly and used as intended. Here are some guidelines:

• **Student course ratings are only one source of evidence regarding teaching effectiveness.** Students are in the best position to judge certain aspects of teaching, such as the effectiveness of lectures, course activities and the instructor's day-to-day behavior. Students can also share perceptions of their learning.

• When making judgments about teaching effectiveness in a personnel review (such as a tenure review), it is best to sample student rating information over several semesters and from a variety of courses taught by the instructor (if the instructor teaches a variety of courses).

• When reviewing student ratings of instruction, first determine how representative the sample is, then look at the results. Finally, review any comparative information.

• **The first three items on the survey provide an overall evaluation of teaching,** much like a final course grade; they provide the best information for evaluating teaching effectiveness. The remaining questions offer student feedback regarding specific teaching behaviors; their primary purpose is to gather information that can be used to improve teaching.

• **(Sample) Size Matters.** In courses with only a handful of students completing the forms, scores become highly sensitive to one or two outliers – the occasional extremely disgruntled, or extremely uncritical, student. In larger classes we can have more confidence that these outliers will average out. Because of this natural variability among raters, SRI scores for small classes must be looked at with care, or grouped together with other classes taught by the same instructor.

• **Response rates are important,** even in larger classes. Unless every student responds, there is always the possibility that those who do may represent a biased sample of the class. In general, response rates of less than sixty-five percent should be viewed with caution.

• **Don’t make mountains out of molehills.** Small numerical differences in ratings are unlikely to be meaningful; for example, there is probably no real difference in teaching effectiveness between an instructor with a mean rating of 4.0 and one with a rating of 3.9 or 3.8. While student ratings of instruction are reliable, the means should not be viewed as precise, absolute numbers, because there is a margin of error, and confidence levels can vary based on response rate and sample size. Teaching evaluations are most
effectively used to evaluate teaching according to broad categories, such as “exceptional,” “acceptable,” or “unacceptable.”

- **Both the raw score and the comparison of means between instructors in the same program provide useful information.** For example, a mean of 4.3 could possibly fall below the mean score of all instructors in a program. In this case the instructor has been rated below the average, but it is important to remember that students have still judged the course positively.

- **The standard deviation is widely used in statistics to show how much variation there is in the responses.** (Standard deviation is included on your reports, labeled as STD.) A low standard deviation (e.g. less than one) indicates that respondents generally agreed while a larger standard deviation indicates a wider range of student responses.