

## Using Student Evaluation of Teaching Surveys

Student evaluations are useful as one component in the overall evaluation of instructional effectiveness. However, high ratings do not guarantee effective instruction, nor do low ratings always mean ineffective instruction. Peer evaluation of instruction, as mandated by the Senate Policy on Evaluation of Teaching Contributions, self evaluations, and reports of teaching activities and instructional development are among other important information sources that should be considered in the evaluation of teaching.

For personnel purposes, a small set of global questions is more appropriate than a larger group of specific questions (e.g., instructor's rapport with students or quality of textbook). The most useful single item is "overall teaching effectiveness." Student ratings of the instructor's knowledge of the subject matter are best interpreted strictly as impressions.

Research suggests that student evaluations are reliable, i.e., average ratings are reasonably consistent and do not change significantly from one offering of a course to another, provided that neither the student composition of the course nor the instructional methods have changed significantly. The response rate is important. Data from classes in which fewer than 75 percent of the students respond or in which there are fewer than 15 students may not provide reliable information.

Student evaluations are a valid measure, but are not synonymous with other measures, of teaching effectiveness. The correlation of global student evaluations with external measures of amount learned (e.g., final examination scores or course grades) is moderate at best.

Avoid using data from a single evaluation of a course. Data collected over time are necessary to verify the consistency of ratings of instruction and provide information about changes in ratings.

A range of student ratings is expected. Typically, ratings are concentrated at the higher end of the rating scale, with a long "tail" consisting of a few low ratings. However, different distributions sometimes may occur and call attention to the possibility of subgroups of students within the course.

Research indicates that average ratings may vary with characteristics (co-variates) such as class size, class format (e.g., lecture or discussion), course level, elective or required, and instructor experience. Ratings tend to be higher for smaller classes, for electives, for classes with a discussion format, and for classes taught by an instructor with more experience. These possible co-variates are important in selecting appropriate norms with which to compare an individual. However, appropriate norms cannot be provided until an adequate data base is accumulated.

---

We gratefully acknowledge the permission from Office of Measurement Services,  
University of Minnesota to post this on the CDTL website.

Student evaluations tend, in general, to be consistent across certain student characteristics. Average evaluations usually do not vary significantly with age, sex, class standing, grade point average, or other indices of student ability. Little research has been done on the relationship between average ratings and the ethnicity of the students rating the course.

Student evaluation data are at least ordinal (rank). Therefore, the median is an appropriate measure of the central rating, and the frequency distribution and associated percentages measure the variability of the ratings. To the extent that the verbal anchors induce respondents to apply uniform differences between response categories, interval statistics such as the mean and the standard deviation are appropriate measures of the central rating and the variability in ratings.

Avoid drawing inferences or conclusions based upon small differences in ratings. Responses are assigned integer values, but statistics are reported with decimal precision. It is tempting but not justifiable to infer differential teaching effectiveness from small differences in average ratings. To obtain statistical significance, differences of 0.5 to 1.5 points may be needed. Moreover, statistically significant differences may not indicate important differences in instructional effectiveness, especially at the higher end of the ratings scale.

The results of your student evaluations of teaching are on the following pages. If there were fewer than 10 questionnaires, only the response counts are given; otherwise, the number and percent of responses to each choice are shown. These distributions and the corresponding graphs show you how your students answered each item.

The mean, median, and standard deviation also are listed for groups of 10 or more.

The mean is the average of the responses to an item based on a 1-7 or 1-5 scale. The values of all responses are added together and divided by the number of responses. The mean is a good measure for summarizing results if the distribution of responses approximates a bell-shaped curve.

The median indicates the mid-point of the item responses--the point at which 50% of the responses are higher and 50% are lower. The median is a better summary than the mean if the distribution is not symmetrical but appears to have a long tail at one end.

Sometimes the mean or median may hide important aspects of the distribution. For example, if 1/2 marked 1 and 1/2 marked 7, the resulting mean and median of 4 would mask an important fact: the variability of the ratings. The standard deviation is an index of variability or consensus among the responses; the lower the standard deviation, the greater the similarity of responses. The highest standard deviations indicate items for which the responses are the most spread out. Generally on a 5 to 7-point scale a standard deviation of 0.8 to 1.0 or less can be considered low.

Checklist items (penultimate page), however, typically do not receive high frequencies of response. Those that do may deserve attention even if the percentage of responses is fairly low.

In response to students' ongoing requests for information, the Senate Committee on Educational Policy appointed a committee of faculty and students to develop a set of ten questions to guide students in selecting courses. The results of these questions will give you an opportunity to review your results, regardless of whether or not you give permission to release results to students. Results are course and term specific and decisions to release these results are only made by instructors.

For additional information about evaluating and improving teaching, visit the Office of Measurement Services homepage at:

<http://www.ucs.umn.edu/oms/>