

Recommendations for Use of Student Ratings of Instructors in Merit and Promotion Decisions

Student course ratings have many uses, particularly if viewed over time and across courses. Student ratings provide information that instructors can use to identify areas of strength and areas needing improvement in their teaching. Students also can use course ratings to help them make enrollment choices. Furthermore, departments and teaching units can use student ratings in the aggregate to assess the overall performance of multi-course and multi-instructor units, as well as to evaluate individual instructors for personnel reasons, such as decisions regarding retention, promotion, tenure and merit pay.

The nine recommendations listed below provide guidelines for the use of student course ratings in personnel decisions.

1. *Student ratings must be used in concert with other data that relate to the quality of a faculty members teaching, rather than as a sole indicator of teaching quality.* Other sources such as peer reviews of classroom sessions, peer reviews of curricular materials, and faculty self-reflection should be assessed in addition to student evaluations to gain a true sense of the teaching skills and performance of a faculty member. Consideration of these other sources of evidence is especially important because student ratings alone do not provide sufficient evidence of the extent of student learning in a course.
2. *Evaluations from more than a single section should be used in making any decision about teaching quality.* Research has shown that ratings from at least five courses are necessary to assure adequate reliability. The validity of the ratings for measuring teaching quality is increased as a greater variety of course formats is represented in the data upon which decisions are based. Trends in ratings across years may also be important in assessing teaching.
3. *Only global ratings of teaching effectiveness, as represented by the first four items on Instructional Assessment forms, should be used in personnel decisions.* Other, more specific items should be used by the faculty member for review of specific skills and areas for improvement.
4. *Small differences in individual evaluations should not be used as a basis for differential decisions.* Because student ratings yield numerical averages, there is a temptation to overestimate the precision of the averages that are presented. Small differences in ratings may not be meaningful. It is better to deal with much broader classifications, such as Excellent vs. Very Good.
5. *Interpretations of student ratings averages should be guided by awareness that, in a university that uses teaching excellence as a hiring criterion, teaching is typically at a high level.* Also, students tend to rate faculty at or near the high end of the scale. Only about 5% of University of Washington faculty receive average ratings below the scale point labeled "Good." It is therefore not appropriate to use the median (or 50th percentile) as a presumed dividing line between strong and weak teachers. More appropriate would be to assume that the great majority of teachers are strong. It is also appropriate, when evaluating average ratings of individual instructors, to consider relevant comparisons (see Recommendation 6) and specific characteristics of courses taught (see Recommendation 7).
6. *Comparative data should be used but with caution.* Colleges and departments should compile information about evaluations for faculty in their courses over time for comparative purposes. However, for comparisons to be useful, the normative group should be based on more than a narrow population of instructors. Smaller departments may not want to rely on departmental norms but use norms calculated for a number of similar departments or for the school or college as a whole. At times, it may be better to compare ratings of similar courses across departments rather than ratings of dissimilar courses within departments.
7. *Course characteristics should be considered when interpreting results.* For example, large lecture courses typically receive lower ratings than smaller courses, new courses being taught for the first time receive lower ratings than well-established courses, introductory courses for non-majors receive lower ratings than higher division courses for majors, and courses in departments that have high student workloads typically receive lower ratings. Adjustments for course type should be made in

order to have a fairer sense of the faculty member's teaching skills. One way to adjust for course types is by choosing similar courses for normative comparisons.

8. *Faculty members should be given an opportunity to respond to evaluation results.* Faculty should have an opportunity to discuss the objectives of the course, how the teaching methods were used to meet that objective, and how circumstances in the course might have affected evaluations. Furthermore, other evaluation information gained from a given course (see Recommendation 1) can aid with the interpretation of ratings results.
9. *Administration of course ratings should be scheduled to maximize the number of respondents.* Generally, evaluations will have greater validity when higher proportions of the enrolled students complete evaluation forms. Ratings may not be an accurate reflection of the entire class when smaller proportions of students respond. This problem can be particularly acute in small classes. Generally, a minimum of a third of enrolled students must be included in the results to have any confidence in the results. As proportions decrease, particularly in small classes, there is greater opportunity for the rating of one or a few students to disproportionately affect the results.

Additional Resources:

The recommendations made in this document are based on research that has been summarized in a number of sources. Three good resources for additional information about student evaluations are:

Braskamp, L.A., & Ory, J.C. (1994). *Assessing Faculty Work: Enhancing Individual and Institutional Performance*. San Francisco: Jossey-Bass.

Lewis, K.G. (2001, Fall). Techniques and Strategies for Interpreting Student Evaluations. *New Directions for Teaching And Learning*, 87. San Francisco: Jossey-Bass.

National Research Council. (2003). *Evaluating and Improving Undergraduate Teaching in Science, Technology, Engineering, and Mathematics*. Washington, D.C.: National Academies Press.

Approved by FCIQ, April 18, 2003