

Interpreting Student Evaluation of Teaching (SET) Results: Guidelines for Deans, Department Heads, and Faculty

Prepared by Faculty Standards Committee (FSC), April 7, 2011

Approved by the University Senate April 25, 2011

In March 2010, the University Senate passed a motion endorsing the use of student evaluations of teaching (SETs), recognizing that they provide information that can be useful for improving teaching and evaluating teaching performance. However, the Senate also urged caution in interpreting numerical values from SETs as an indicator of teaching competence. This caution is based on two premises. First, as explicitly recognized by the Senate, no set of numerical values can be sufficient as the sole indicator of teaching effectiveness. Second, while an overall score on an individual teaching evaluation can be an indicator of teaching performance, research shows that SET results are sometimes influenced by factors other than teaching performance, such as student bias.

When used in performance evaluations, SETs can have significant consequences for the careers of both full-time and part-time instructors. Thus, it is imperative that they be interpreted carefully. Toward this end, the University Senate asked the Faculty Standards Committee to develop guidelines for appropriate use of teaching evaluations. The guidelines below were developed by the FSC.

Overall recommendation: In addition to considering the information provided by SETs, Deans, Department Heads, and faculty (including PTR committees) are encouraged to explore other methods of evaluating instructors. To improve instruction, Department Heads should review evidence of teaching performance with the instructor and provide feedback.

Factors other than teaching competence that can influence SET results: Although the literature on SETs is both extensive and complicated (see the appendix for a partial list of references) and it is difficult to isolate contributing factors, research suggests that SETs are sometimes influenced by the following factors:

- *Student year:* First-year students tend to give the lowest ratings; graduate students the highest.
- *Course-level:* Students tend to give lower ratings in required courses than in electives.
- *Instructor ethnicity:* Students sometimes give faculty of color lower ratings.
- *English as a second language:* Students sometimes give lower ratings to instructors who speak English as a second language.
- *Discipline:* Students sometimes give lower ratings to women in male-dominated disciplines such as science, mathematics, economics, engineering, and philosophy, or to men in female-dominated disciplines such as nursing.
- *Gender:* Students can have different expectations for male and female teachers and sometimes give higher ratings to members of their own gender.
- *Field of study/discipline:* The balance of research evidence shows that classes in sciences and engineering tend to receive lower ratings than those in the humanities.

Note: While some anecdotal evidence and popular belief might suggest that SET results are correlated with expected grades (with easy graders receiving higher scores), this claim is not supported by systematic research. Rather, evidence shows that there is a strong correlation between instructor ratings and students' perception of learning outcomes.

Guidelines for Interpreting SET results: Based on research related to SETs, the Senate recommends the following guidelines be used in interpreting SET results.

1. Look for patterns over time. Compare multiple courses across multiple semesters to form generalizations about teaching effectiveness.
2. Remember that the sample is not random and therefore may not be representative of the entire class
3. Do not over-interpret small differences in median ratings.
4. Do not use university-norm results as a line separating "failing" and "passing" teaching performance.
5. Do not average multiple, inherently-different SET items into a single value.
6. Ask: Are one or two low student ratings affecting the results in a small class?
7. Ask: Does this instructor receive consistently better ratings for some skills than others (preparation, clear assignments, receptivity to students)?
8. Ask: Are SET ratings influenced by large class size or courses outside of a student's major?
9. Ask: Are SET ratings in particular classes bi-modal, as sometimes occurs in classes that include controversial or politically-charged topics?
10. Focus on the two questions related to overall ratings of the instructor's teaching and the course.
11. Recognize that when there are responses from small numbers of students, percentages may not be meaningful.

Appendix

Selected Relevant Publications

- Anderson, K. J., & Smith, G. (2005). Students' preconceptions of professors: Benefits and barriers according to ethnicity and gender. *Hispanic Journal of Behavioral Sciences, 27*, (2), 184-201.
- Arreola, R. A. (2007). *Developing a Comprehensive Faculty Evaluation System*. San Francisco: Jossey-Bass.
- Cashin, W. (1999). Student Rating of Teaching: Uses and Misuses. In P. Seldin (Ed.), *Changing Practices in Evaluating Teaching: A Practical Guide to Faculty Performance and Promotion/Tenure Decisions* (pp. 25-44). Boston: Anker Publishing Company, Inc.
- Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *Journal of Higher Education, 71*, p. 17.
- Cohen, P.A. (1990). Bring research into practice. In M. Theall, & J. Franklin (Eds.), *Student Ratings of instruction: Issues for Improving Practice: New Directions for Teaching and Learning*, No. 43 (pp. 123-132). San Francisco: Jossey-Bass.
- Feldman, K. A. (1993). College students' views of male and female faculty college teachers: Part II – Evidence from students' evaluations of their classroom teachers. *Research in Higher Education, 34*, 151-211.
- Hendrix, K. G. (1998). Student perceptions of the influence of race on professor credibility. *Journal of Black Studies, 28*, 738-764.
- Houston, T. Empirical Research on the Impact of Race & Gender in the Evaluation of Teaching. Report, Center for Excellence in Teaching & Learning. Seattle University. October 5, 2005
- Houston, T. Research Report: Race and Gender Bias in Student Evaluations of Teaching. Center for Excellence in Teaching & Learning. Seattle University. October 31, 2005
- Ory, J.C. & Ryan, K. (2001). How do student ratings measure up to a new validity framework? In M. Theall, P.C. Abrami, & L.A. Metz (Eds.), *New Directions for institutional research: No. 109. The student ratings debate: Are they valid?* San Francisco: Jossey-Bass.
- Rubin, D. L. (1998). Help! My professor (or doctor or boss) doesn't talk English. In J. N. Martin, T. K. Nakayama, L. A. Flores (Eds.), *Readings in Cultural Contexts* (pp. 149 – 160). Mountain View, CA: Mayfield Publishing Company.
- Seldin, P. (1999) *Changing practices in Evaluating Teaching*. Bolton, MA. Anker.