

## Report to the Faculty Senate

From: Ad Hoc Committee on Student Evaluations of Faculty Teaching

Date: April 10, 2024

The committee has reviewed relevant literature, SU's current practices, and considered best practices with respect to the role of Student Evaluations of Teaching (hereafter SET) for evaluating faculty for Tenure and Promotion. The evidence is that SET provide little, if any, valuable information about faculty teaching and student learning because SET register wide range of student biases regarding discipline, race, gender, physical attractiveness, likeability, and ease of grading. (See appendices for details and an extensive review of relevant literature).

Below are our recommendations.

1. Faculty are expected to solicit student feedback on classroom experiences, particularly in the form of mid-semester formative evaluations, and use that information to guide adjustments in their courses. (See Appendix B).
2. Faculty are not expected/required to share the raw data of SET with supervisors or as part of Tenure or Promotion packets.
  - a. Faculty must continue to report on their responses to SET in their annual self-evaluation; inclusion of raw data in those reports is at the discretion of the individual faculty member.
  - b. Departments and schools are encouraged to use other practices for evaluating teaching that serve their programs, for example peer evaluations. GULLWEEK, as well as our courses, allow us to assess student learning.
3. Faculty will develop and maintain a teaching portfolio for the purposes of evaluation of their teaching for Tenure and Promotion.
  - a. Teaching Portfolios will include relevant teaching materials, such as syllabi, assignments, sample student work, and other, as well as self-evaluation materials.
  - b. Departments/schools will develop guidelines for teaching portfolio contents and structure (See Appendix C)
4. Full-time, tenure-track, faculty will participate in ongoing, small, interdisciplinary support groups for faculty development in teaching excellence as part of their progress toward

with the University's mission and General Education Program, and desire to support junior faculty.

- b. Optimal size for groups is no larger than 6 (7 with facilitator).
- 5.

## **Appendix A: SET Literature Summary**

*For every complex problem there is an answer that is*

- A more recent meta-analysis of nearly 100 multi-section studies indicates that SET /learning correlation is small ( $r = 0.12$ ). When prior student ability is considered, the correlation is zero ( $r = -0.06$ ). (Uttle, 2017)
- A 2016 study of 23,000 SET scores from 4,423 first year students in 1,177 sections in France found the correlation between SET and final exam scores to be  $r = 0.04$ . Of note, SET were compulsory, so the student response rate was nearly 100%, and the students had been unable to self-select into different sections. (Boring, 2016).

SET are consistent, in that evaluations for a given instructor positively correlate within the same course and over time (Carpenter, 2020). This suggests that they reflect something stable about the instructor. Research points to a number of factors, including:

**Course Discipline** – Multiple studies





Basow, S.A. and Martin, J.L. (2012) Bias in Student Evaluations. In: Kite, M.E., Ed., *Effective Evaluation of Teaching: A Guide for Faculty and Administrators*, Society for the Teaching of Psychology, Washington DC, 40-49.

Boring, A., Ottoboni, K., & Stark, P. S. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*, 0(0):1-11.  
<http://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1>

Braga, M., Paccagnella, M., & Pellizzari, M. Evaluating









ones, “ which use [SE] evidence to ‘sum up’ our overall performance or status to decide about our annual merit pay, promotion, and tenure. However, SES is the single type evidence the SU faculty handbook requires faculty to include in a tenure and promotion dossier. In addition,

## II. Rethinking Faculty Evaluation Methods

Simonson, Earl,

Like Simonson, Earl, and Frary, several other researchers asserted that the demonstration and discussion of learning outcomes might provide more accurate information regarding a faculty member's teaching effectiveness. To this end, Anders proposed using **focus groups and role-play** to solicit more candid and detailed reflections from students about their learning in a course. Borch, Sandvoll, and Risor advocated a similar type of tool by suggesting that faculty collaborate with students to create **-based evaluation**. However, the Borch, Sandvoll, and Risor researched was conducted in Norway, and their proposed method raises important questions regarding the resources needed to execute it. Lastly, Stark-Wroblewski, Ahlering, and Brill, suggest that faculty conduct **pre- and post-knowledge of a course-related topic** to measure student learning.

### III. Peer Observations Concerns

Although several of the aforementioned studies assert that peer statements/observations and class room visits can help mitigate the bias and other problems with student evaluations, Berk reveals that most faculty are resistant to them because their potential for bias, unfairness, and inaccuracy. Indeed, there is "consensus" in academia that "peer observation data should be used for formative [or developmental] rather than summative decisions." Yet, many departments at SU require faculty to include peer observations in their tenure and promotion application. Indeed, J.M. Golding and Philipp Kraemer question whether peer observations can infringe on academic freedom; therefore, our subcommittee might request that the Senate Academic Freedom and Tenure Committee explore this concern.

#### SET Design Best Practices

Given the many biases in student evaluation and often flawed interpretation of it, some studies provided suggestions for improving teaching evaluation design, such as 1) dropping the questions beyond students' capability; 2) drop the obscure questions, such as overall teaching effectiveness; 3) avoid comparing averages of teaching evaluation scores; 4) avoid comparing different courses at different course level and features (e.g., Hornstein, 2017). Specifically, Carpenter, Witherby, and Tauber (2020) propose to **develop a well-designed student evaluation** to mitigate the biases: 1) eliminating the evaluation questions beyond students' knowledge and capability (e.g., evaluate the professor's knowledge in the field of study); 2) relying more on students' qualitative comments but the information users need to get trained when interpreting

this information; 3) completing evaluations at multiple times throughout the semester to limit the negative effect of faulty memory by the end of semester. **They also admit that all these methods may not be able to solve the biases with student evaluations.** They suggest the following alternatives: peer evaluation/observation; student interview by administrators; teaching portfolio including one's teaching philosophy, syllabi, example lessons, assignments, and grading rubric; follow-up assessment about students'

comments but the information users need to get trained when interpreting this information; 3) completing evaluations at multiple times throughout the semester to limit the negative effect of faulty memory by the end of semester.

### **Bibliography**

- Anders, B. A. (2023, May). Using a Focus Group to Enhance Course Evaluation Inclusion and Feedback. In *Erasmus Scientific Days 2022 (ESD 2022)* (pp. 332-338). Atlantis Press.
- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International journal of teaching and learning in higher education*, 17(1), 48-62.
- Bork, I., Sandvoll, R. & Risor T. (2020). Discrepancies in purposes of student course evaluations: what does it mean to be 'satisfied'? *Educational Assessment, Evaluation and Accountability*, 32, 83-102.
- Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020). On students' (mis) judgments of learning and teaching effectiveness. *Journal of Applied research in memory and cognition*, 9(2), 137-151.
- Golding, J.M. & Kraemer, P.J. Observations of Professors: Tread Lightly. *Inside Higher Education*. July 23 2015. Website.
- Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1), 1304016.
- Miller, J. E., & Seldin, P. (2014). Changing practices in faculty evaluation. *Academe*, 100(3), 35-38.
- Simonson, S. R., Earl, B., & Frary, M. (2022). Establishing a framework for assessing teaching effectiveness. *College Teaching*, 70(2), 164-180.
- Shao, L. P., Anderson, L. P., & Newsome, M. (2007). Evaluating teaching effectiveness: Where we are and where we should be. *Assessment & Evaluation in Higher Education*, 32(3), 355-371.
- Stark Wroblewski, K., Ahlering, R. F., & Brill, F. M. (2007). Toward a more comprehensive approach to evaluating teaching effectiveness: Supplementing student evaluations of





[Teaching Portfolios | Center for Teaching | Vanderbilt University](#)

Colorado State University (University of Dayton Teaching Portfolio Guidelines)

[University-of-Dayton-Teaching-Portfolio-Guide.pdf \(colostate.edu\)](#)

