Statement on Student Evaluations of Teaching

American Sociological Association

September 2019

Most faculty in North America are evaluated, in part, on their teaching effectiveness. This is typically measured with student evaluations of teaching (SETs), instruments that ask students to rate instructors on a series of mostly closed-ended items. Because these instruments are cheap, easy to implement, and provide a simple way to gather information, they are the most common method used to evaluate faculty teaching for hiring, tenure, promotion, contract renewal, and merit raises.

Despite the ubiquity of SETs, a growing body of evidence suggests that their use in personnel decisions is problematic. SETs are weakly related to other measures of teaching effectiveness and student learning (Boring, Ottoboni, and Stark 2016; Uttl, White, and Gonzalez 2017); they are used in statistically problematic ways (e.g., categorical measures are treated as interval, response rates are ignored, small differences are given undue weight, and distributions are not reported) (Boysen 2015; Stark and Freishtat 2014); and they can be influenced by course characteristics like time of day, subject, class size, and whether the course is required, all of which are unrelated to teaching effectiveness.

In addition, in both observational studies and experiments, SETs have been found to be biased against women and people of color (for recent reviews of the literature, see Basow and Martin 2012 and Spooren, Brockx, and Mortelmans 2015). For example, students rate women instructors lower than they rate men, even when they exhibit the same teaching behaviors (Boring, Ottoboni, and Stark 2016; MacNell, Driscol, and Hunt 2015), and students use stereotypically gendered language in how they evaluate their instructors (Mitchell and Martin 2018). The instrument design can also affect gender bias in evaluations; in an article in *American Sociological Review*, Rivera and Tilcsik (2019) find that the range of the rating scale

(e.g., a 6-point scale versus a 10-point scale) can affect how women are evaluated relative to men in male-dominated fields. Further, Black and Asian faculty members are evaluated less positively than White faculty (Bavishi, Madera, and Hebl 2010; Reid 2010; Smith and Hawkins 2011), especially by students who are White men. Faculty ethnicity and gender also mediate how students rate instructor characteristics like leniency and warmth (Anderson and Smith 2005).

A scholarly consensus has emerged that using SETs as the primary measure of teaching effectiveness in faculty review processes can systematically disadvantage faculty from marginalized groups. This can be especially consequential for contingent faculty for whom a small difference in average scores can mean the difference between contract renewal and dismissal.

Given these limitations, the American Sociological Association, in collaboration with the scholarly societies listed below, encourages institutions to use evidence-based best practices for collecting and using student feedback about teaching (Barre 2015; Dennin et al. 2017; Linse 2017; Stark and Freishtat 2014). These include:

1. Questions on SETs should focus on student experiences, and the instruments should be framed as an opportunity for student feedback, rather than an opportunity for formal ratings of teaching effectiveness. For example, two universities – Augsburg University and University of North Carolina Asheville – recently revised and renamed their instruments to the "University Course Survey" and the "Student Feedback on Instruction Form," respectively, to emphasize that student feedback, while important, is not an evaluation of teaching effectiveness.

2. SETs should not be used as the only evidence of teaching effectiveness. Rather, when they are used, they should be part of a holistic assessment that includes peer observations, reviews of teaching materials, and instructor self-reflections. This holistic approach has been in wide use at teaching-focused institutions for many years and is becoming more common at research institutions as well. For example:

   - University of Oregon has undertaken a multi-year process to develop a holistic framework for assessing teaching effectiveness, including peer review, self-reflection, and student feedback. Extensive research and resources are available on the Office of the Provost website, including guidance on how to interpret SETs

   - University of Southern California has instituted peer review of teaching for faculty evaluation. Their Center for Excellence in Teaching provides resources for how to use peer review effectively and addresses common concerns.

   - University of California Irvine requires faculty to submit two types of evidence to document teaching effectiveness. In addition to SETs, faculty can submit a reflective teaching statement, peer evaluations of teaching, and other evidence like a Teaching Practices Inventory, developed by physicist Carl Weiman.

   - University of Nebraska Lincoln has articulated best practices for faculty evaluation that state, in part, "it is recommended that student evaluation scores should not be given undue weight in faculty evaluations, since these scores are easily manipulated and reflect many attitudes that extend beyond the successful accomplishment of the faculty member's teaching duties."

   - The University of Michigan's Center for Research on Teaching and Learning recommends that student ratings should never be used in isolation and should be part of a broader assessment of teaching effectiveness. They have developed resources that include a summary of research findings on SETs, a handout for students on how to make their feedback most helpful to instructors, and best practices for using SETs in personnel decisions.

   - Ryerson University has gone even further and is no longer using SETs for tenure or promotion decisions (Farr 2018). Instead, Ryerson asks faculty to compile a teaching dossier that includes a statement of teaching philosophy, evidence of curricular engagement, and self-reflections.

3. SETs should not be used to compare individual faculty members to each other or to a department average. As part of a holistic assessment, they can appropriately be used to document patterns in an instructor's feedback over time.

4. If quantitative scores are reported, they should include distributions, sample sizes, and response rates for each question on the instrument (Stark and Freishtat 2014). This provides an interpretative context for the scores (e.g., items with low response rates should be given little weight).

5. Evaluators (e.g., chairs, deans, hiring committees, tenure and promotion committees) should be trained in how to interpret and use SETs as part of a holistic assessment of teaching effectiveness (see Linse 2017 for specific guidance).

Gathering student feedback on their experiences in the classroom is an important part of student-centered teaching. This feedback can help instructors to refine their pedagogies and improve student learning in their courses. However, student feedback should not be used alone as a measure of teaching quality. If it is used in faculty evaluation processes, it should be considered as part of a holistic assessment of teaching effectiveness.

Endorsements

American Anthropological Association

American Dialect Society

American Folklore Society

American Historical Association

American Political Science Association

Association for Slavic, East European, and Eurasian Studies

Dance Studies Association

International Center of Medieval Art

Latin American Studies Association

Middle East Studies Association

National Communication Association

National Council on Public History

Rhetoric Society of America

Society for Cinema and Media Studies

Society for Personality and Social Psychology

Society of Architectural Historians

Sociologists for Women in Society

References

Anderson, Kristin J., and Gabriel Smith. 2005. "Students' Preconceptions of Professors: Benefits and Barriers According to Ethnicity and Gender." *Hispanic Journal of Behavioral Sciences* 27(2):184-20.

Barre, Elizabeth. 2015. "Student Ratings of Instruction: A Literature Review." Reflections on Teaching and Learning: The CTE Blog. Rice University Center for Teaching Excellence. Retrieved from https://cte.rice.edu/blogarchive/2015/02/01/studentratings.

Basow, Susan A., and Julie L. Martin. 2012. "Bias in Student Evaluations." Pp. 40-49 in *Effective Evaluation of Teaching: A Guide for Faculty and Administrators*, edited by Mary E. Kite. Washington, DC: Society for the Teaching of Psychology. Retrieved from http://teachpsych.org/ebooks/evals2012/index.php.

Bavishi, Anish, Juan M. Madera, and Michelle R. Hebl. 2010. "The Effect of Professor Ethnicity and Gender on Student Evaluations: Judged Before Met." *Journal of Diversity in Higher Education* 3(4):245-256.

Boring, Anne, Kellie Ottoboni, and Philip B. Stark. 2016. "Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness." *ScienceOpen Research* (DOI: 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1).

Boysen, Guy A. 2015. "Significant Interpretation of Small Mean Differences in Student Evaluations of Teaching Despite Explicit Warning to Avoid Overinterpretation." *Scholarship of Teaching and Learning in Psychology* 1(2):150-162.

Dennin, Michael, Zachary D. Schultz, Andrew Feig, Noah Finkelstein, Andrea Follmer Greenhoot, Michel Hildreth, Adama K. Leibovich, James D. Martin, Mark B. Moldwin, Diane K. O'Dowd, Lynmarie A. Posey, Tobin L. Smith, and Emily R. Miller. 2017. "Aligning Practice to Policies: Changing the Culture to Recognize and Reward Teaching at Research Institutions." *CBE—Life Sciences Education* 16(5):1-8.

Farr, Moira. 2018. "Arbitration Decision on Student Evaluations of Teaching Applauded by Faculty." *University Affairs* August 28. Retrieved from https://www.universityaffairs.ca/news/news-article/arbitration-decision-on-student-evaluations-of-teaching-applauded-by-faculty/.

Linse, Angela R. 2017. "Interpreting and Using Student Ratings Data: Guidance for Faculty Serving as Administrators and on Evaluation Committees." *Students in Educational Evaluation* 54:94-106.

MacNell, Lillian, Adam Driscoll, and Andrea N. Hunt. 2015. "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching." *Innovative Higher Education* 40(4):291-303.

Mitchell, Kristina M. W., and Jonathan Martin. 2018. "Gender Bias in Student Evaluations." *PS: Political Science and Politics* 51(3):648-652.

Reid, Landon D. 2010. "The Role of Perceived Race and Gender in the Evaluation of College Teaching on RateMyProfessors.com." *Journal*

of Diversity in Higher Education 3(3):137-152.

Rivera, Lauren A., and András Tilcsik. 2019. "Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation." *American Sociological Review* 84(2):248–274.

Smith, Bettye P., and Billy Hawkins. 2011. "Examining Student Evaluations of Black College Faculty: Does Race Matter?" *The Journal of Negro Education* 80(2):149-162.

Spooren, Pieter, Bert Brockx, and Dimitri Mortelmans. 2013. "On the Validity of Student Evaluation of Teaching: The State of the Art." *Review of Educational Research* 83(4):598–642.

Stark, Philip B., and Richard Freishtat. 2014. "An Evaluation of Course Evaluations." *ScienceOpen Research* (DOI: 10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1).

Uttl, Bob, Carmela A. White, and Daniela Wong Gonzalez. 2017. "Meta-Analysis of Faculty's Teaching Effectiveness: Student Evaluation of Teaching Ratings and Student Learning Are Not Related." *Studies in Educational Evaluation* 54(1):22-42.

Additional Resources

Association of American Universities. n.d. *Aligning Practice to Policies: Changing the Culture to Recognize and Reward Teaching at Research Universities.* Retrieved from https://www.aau.edu/sites/default/files/AAU-Files/STEM-Education-Initiative/Aligning-Practice-To-Policies-Digital.pdf

American Educational Research Association. 2013. *Rethinking Faculty Evaluation: AERA Report and Recommendations on Evaluating Education Research, Scholarship, and Teaching in Postsecondary Education.* Retrieved from http://www.aera.net/Portals/38/docs/Education_Research_and_Research_Policy/RethinkingFacultyEval_R4.pdf.

Center for the Integration of Research, Teaching, and Learning (CIRTL). 2018. Local CIRTL Program Evaluation Resource Guide.
Michigan State University. Retrieved from https://www.cirtl.net/resources/708.

Kite, Mary E., ed. 2012. *Effective Evaluation of Teaching: A Guide for Faculty and Administrators.* Washington, DC: Society for the Teaching of Psychology. Retrieved from http://teachpsych.org/ebooks/evals2012/index.php.

Stark, Philip B. 2018. "Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness." Presentation given as part of the Teaching Assessment Working Group Speaker Series, Valuing Teaching: Challenges and Strategies to Value Effective Teaching. Simon Fraser University. Retrieved from https://www.youtube.com/watch?v=5haOjlfJDb8&feature=youtu.be.

Van Valey, Thomas L., ed. 2011. *Peer Review of Teaching: Lessons from and for Departments of Sociology.* Washington, DC: American Sociological Association.

Vasey, Craig, and Linda Carroll. 2016. "How Do We Evaluate Teaching? Findings from a Survey of Faculty Members." *Academe*, May-June. Retrieved from https://www.aaup.org/article/how-do-we-evaluate-teaching.