

A Summary of Resources and Research on Using and Interpreting Teaching Surveys

by Lindsay Onufer, Program Manager, Assessment of Teaching Initiative

Resources

[Tips for Improving Student Feedback and Mitigating the Effects of Biases in Teaching Surveys](#)

[Understanding Your OMET Results](#)

[A Quick Guide for Interpreting OMET Results for Faculty and Administrators](#)

Use of Student Opinion of Teaching Surveys

As in other higher education institutions, nearly all academic units at Pitt use teaching survey data for formative and summative assessment of teaching effectiveness. Because teaching surveys portray the valuable but nonexpert student perspective, experts on assessment of teaching best practices recommend that teaching survey data be reviewed as one of several data sources (Benton & Cashin, 2014; Benton & Li, 2015; Berk, 2005; Burdsal & Harrison, 2008; Gormally, Evans & Brickman, 2014; Linse, 2017; Nasser & Fresko, 2002). Research on assessment of teaching also indicates assessment of teaching, particularly summative evaluations, should require multiple data sources and allow for triangulation of data (Benton & Cashin, 2014; Benton & Li, 2015; Berk, 2005; Burdsal & Harrison, 2008; Gormally, Evans & Brickman, 2014; Linse, 2017; Nasser & Fresko, 2002). Results of faculty surveys indicate that this is consistent with faculty preferences (Nasser & Fresko, 2002; Vasey & Carroll, 2016). One reason why teaching surveys should be used as one of several data sources is that, like other types of assessment, their results may be influenced by bias.

Bias and Teaching Surveys

Several recent studies (Fan et al., 2019; Mengel, Sauermann, & Zölitz, 2019; Mitchell & Martin, 2018; Peterson, Biederman, Andersen, Ditonto, & Roe, 2019) have yielded results affirming that students' (particularly male students') gender bias had a negative influence on their ratings of female instructors. Peterson, Biederman, Andersen, Ditonto, and Roe (2019) examined whether adding language to raise students' awareness of their own implicit gender bias and to encourage them to rate faculty based on quality of teaching, would lower evidence of gender bias in teaching surveys in two pairs of large introductory biology and American politics courses at Iowa State University to conduct the experiment. In each pair of courses, one female instructor and one male instructor (all white) taught a section. Controlling for portion of students of each gender, mean student level of the class, mean anticipated grade in the class, and mean GPA between the treatment and control groups, students were randomly selected to receive an online teaching survey either with or without language explaining that teaching surveys are subject to implicit bias and to avoid bias and stereotyping when completing the form. The researchers then examined the results of three questions which asked students to rate the instructor, teaching effectiveness, and overall course quality. Ratings for female instructors from students in the treatment group were higher (between .30 and .51 points on a five-point scale) than those for the control group. There was no difference in scores for male instructors in the treatment vs. control groups. Researchers also found that the treatment had no effect on female students' ratings of

female instructors but did influence male students' ratings of female instructors. In both cases the difference was statistically significant for questions about the instructor and overall course quality, but not the question about teaching effectiveness. However, post hoc tests revealed a lack of statistically significant variations in ratings based on student gender. Findings indicate either a slight negative male student bias against female instructors or bias against female instructors unrelated to student gender which was to some extent mitigated by the intervention.

Similarly, two of the largest studies (Fan et al., 2019; Mengel, Sauermann, & Zölitz, 2019) investigating gender bias (and, in the case of the former, cultural bias) in teaching survey results discovered negative bias against female instructors. Fan et al. (2019) used teaching survey data from all schools/departments of a large, public Australian university collected over seven years, examining over a half a million teaching survey results for over 2000 courses and over 3000 instructors. Mengel, Sauermann, and Zölitz (2019) examined two years of undergraduate and graduate student teaching surveys consisting of results from nearly 20,000 teaching survey results from the School of Business and Economics of Maastricht University in the Netherlands. In addition to finding that female instructors received lower ratings, Fan et al. (2019) also discovered that instructors from different cultural backgrounds and non-native English-speaking instructors received lower ratings than instructors from the dominant culture and native English speakers. Differences were most significant for female non-native English-speaking instructors and minoritized instructors in science and business disciplines. Across almost all disciplines, English-speaking male instructors were likeliest to receive the highest rating available. As in other studies (Peterson, Biederman, Andersen, Ditonto, & Roe, 2019), Fan et al. (2019) found that students rated instructors of like gender higher, but also discovered that students rated instructors of similar cultural backgrounds higher. Though Mengel, Sauermann, & Zölitz's (2019) findings were consistent with those of other studies in that male students rated female instructors (approximately 0.2 points on a five-point scale lower) than they did male instructors, interestingly, their results also indicated that female students also rated female instructors lower than male students, though less so. Although the difference in ratings, fractions of a single point on a five-point scale, seems insignificant, the researchers found that it was large enough difference to significantly affect instructor ranking, which would affect career prospects (Mengel, Sauermann, & Zölitz, 2019).

Mitchell and Martin (2018) found similarly noteworthy evidence of gender bias. As in the aforementioned studies, they conducted a quantitative analysis of teaching survey results of a male and female instructor to look for evidence of gender bias, however, their research was unique in that they used teaching surveys from online courses, allowing for more control in keeping course content and delivery similar. They also completed a content analysis of student comments on teaching survey open-ended questions and on student comments on ratemyprofessor.com, a website on which students can leave ratings and written reviews of instructors. In their analysis of Likert scale questions, the researchers divided questions by type into five categories: questions about the instructor, the instructor/course, course, technology, and administrative questions pertaining to procedures like registration and advising. They found that the female instructor was rated between 0.19 and 0.4 points lower than the male instructor in all categories except the one related to administrative procedures, in which they were rated nearly identically.

Their conclusion that differences in ratings suggested a correlation between the gender of the instructor and teaching survey scores was reinforced by student comment content analysis which

revealed that, in a comparison between comments received for a male vs. a female instructor, students were more likely to comment on the female instructor's personality (11.2% more often), how entertaining she was (17% more often) and were more likely (9.2% more often) to refer to her as "teacher" than as "professor" (17.1% less often). Like Mitchell and Martin (2018), Storage, Horne, Cimpian, and Leslie (2016) also used ratings from *ratemyprofessor.com* and found evidence of gender bias. Specifically, they sought to identify differences in the frequency of words like "brilliance" and "genius" in student comments by instructor gender. The researchers theorized that, due to predominant stereotypes that women are less likely to display brilliance than men, comments referencing those traits would appear less frequently in ratings for female instructors. For male vs. female instructors, they found that the word "brilliant" was used at a 1.81:1 ratio and "genius" at a 3.10:1 ratio, meaning that those terms were applied to male instructors significantly more often than they were to female instructors. Interestingly, ratios related to the words "excellent" and "amazing" were similar for instructors of both genders, suggesting that differences in ratings were not just more negative for women, but gender-biased when it came to terms referencing intellect. Although it is possible that the male instructors were simply the better teachers in both students, the quantitative analysis combined with content analysis of student comments strongly suggests that gender bias affected student feedback. Findings of these studies demonstrate that bias affects teaching survey results, but how bias shapes ratings and the significance of its effect varies.

Although, as noted in reviews of research on teaching surveys (Benton & Cashin, 2014; Linse, 2017), there is less research on racial bias in teaching surveys, the studies that address this issue and demonstrate that, like gender bias, there is evidence of a correlation between instructor race and teaching survey ratings. In a 2011 study, Smith and Hawkins examined data from 13,702 teaching surveys collected from undergraduates at a research-intensive university over three years. On the 28-question teaching survey used, white instructors were rated significantly, both statistically and in effect size, higher than Black instructors on every single item. Ratings for instructors who identified as a race other than Black or white generally received similar or slightly higher ratings than white instructors. A later study conducted at the same institution (Smith & Johnson-Bailey, 2012) specifically focused on teaching surveys of female instructors of various races and produced similar results: white female faculty received significantly higher scores on all 28 teaching survey questions.

Taken together, the results of these studies make a compelling argument that student ratings of female instructors, instructors of color (particularly Black instructors), and instructors of different cultural backgrounds than the students they teach are affected by student biases. While they affirm previous assertions that prevalence and significance of bias differ in different contexts (Benton & Cashin, 2014; Linse, 2017), they also suggest that evidence of bias in teaching surveys is growing.

References on the Use of and Bias in Teaching Surveys

- Benton, S.L. & Cashin, W.E. (2014). Student ratings of teaching: A summary of research and literature. IDEA Paper #50. *IDEA*.
https://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/PaperIDEA_50.pdf
- Benton, S.L. & Li, D. (2015). Response to “A better way to evaluate undergraduate teaching.” IDEA Editorial Note #1. *IDEA*.
https://www.ideaedu.org/Portals/0/Uploads/Documents/A_Better_Way_to_Evaluate.pdf
- Berk, R.A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1).
<http://www.isetl.org/ijtlhe/pdf/IJTLHE8.pdf>
- Burdsal, C.A. & Harrison, P.D. (2008). Further evidence supporting the validity of both a multidimensional profile and an overall evaluation of teaching effectiveness. *Assessment & Evaluation in Higher Education*, 33(5), 567-576. doi: 10.1080/02602930701699049
- Fan, Y., Shepherd, L. J., Slavich, E., Waters, D., Stone, M., Abel, R., & Johnston, E. L. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PloS One*, 14(2). doi:10.1371/journal.pone.0209749
- Gormally, C., Evans, M, & Brickman, P. (2014). Feedback about teaching in higher ed: Neglected opportunities to promote change. *CBE-Life Sciences Education*, 13, 187-199. doi: 10.1187/CBE.13-12-0235
- Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, 94-106. doi:10.1016/j.stueduc.2016.12.004
- Mengel, F., Sauermaun, J., Zölit, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2), 535-566. doi:10.1093/jeea/jvx057
- Mitchell, K. M. W., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, 51(3), 648-652. doi:10.1017/S104909651800001X
- Nasser, F. & Fresko, B. (2002) Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education*, 27(2), 187-198. doi: 10.1080/02602930220128751
- Peterson, D.A., Biederman, L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PloS One*, 14(5). doi:10.1371/journal.pone.0216241
- Punyanunt-Carter, N. & Carter, S. L. (2015). Students' gender bias in teaching evaluations. *Higher Learning Research Communications*, 5(3), 28. doi:10.18870/hlrc.v5i3.234
- Smith, B. P., & Hawkins, B. (2011). Examining student evaluations of black college faculty: Does race matter? *The Journal of Negro Education*, 80(2), 149–162.
<https://www.jstor.org/stable/41341117>
- Smith, B. P., & Johnson-Bailey, J. (2012). Student ratings of teaching effectiveness: Implications for non-white women in the academy. *Negro Educational Review*, 62-63(1-4), 115.

- Storage, D., Horne, Z., Cimpian, A., & Leslie, S. (2016). The frequency of "brilliant" and "genius" in teaching evaluations predicts the representation of women and African Americans across fields. *PLoS One*, *11*(3), e0150194. doi:10.1371/journal.pone.0150194
- Vasey, C. & Carroll, L. (2016). How do we evaluate teaching? Findings from a survey of faculty members. *Academe*.
<https://www.aaup.org/article/how-do-we-evaluate-teaching#.XRokwjYpaR>