CLE Committee Report on Biases in Student Evaluation of Teaching

Bias in course evaluations is likely, since by their very nature, course evaluations constitute a voluntary response sample, where the results are usually not representative of the population but pertain mostly to those who have strong feelings or opinions about the topic at hand. However, in addition to the regular bias one would expect in course learning evaluations, many studies have shown that there is additional bias toward women and minority groups in academia. "Even when other factors such as individual course variation, individual teacher variation, student average score, course type and so on are accounted for, gender and culture are found to be statistically significant in some faculties, especially in Science and Business…the most affected group: female instructors from non-English speaking backgrounds…In all faculties, a statistically significant effect against them is observed, with the effect stronger among local students." (Fan et al. 2019:6).

Another study showed that "the language students use in evaluations regarding male professors is significantly different than language used in evaluating female professors. They also show that a male instructor administering an identical online course as a female instructor receives higher ordinal scores in teaching evaluations, even when questions are not instructor-specific." (Mitchell et al., 2018) In an experiment, "assistant instructors in an online class each operated under two different gender identities. Students rated the male identity significantly higher than the female identity, regardless of the instructor's actual gender, demonstrating gender bias." (MacNell et al., 2015).

A study using RateMyProfessors.com "showed that racial minority faculty, particularly Blacks and Asians, were evaluated more negatively than White faculty in terms of overall quality, helpfulness, and clarity, but were rated higher on easiness." (Reid, 2010). One study even examined the effects of professors' race and clothing style on student evaluations. And found that "Black professors have a significant disadvantage in the student evaluation process when compared to White professors." (Aruguete et al., 2017).

In considering biases that exist within and impact student evaluations of teaching (SET), Gutierrez y Muhs et al. (2012) believe that students' evaluations may reflect their own biases rather than the competency of their instructors. They support the need to use a comprehensive

teaching portfolio, which reflects teaching, student advisement, and consulting, for evaluating faculty rather than relying on end-of-course teaching evaluation for appraising faculty performance. In addition, they emphasize the importance of using consistent evaluation for all faculty members.

The book by Gutierrez y Muhs (2012) also discusses the biases and challenges experienced by women of color in academia. The authors discuss the loneliness, lack of support, and racism that these faculty of color face within higher education. They also support the need for white faculty and administrations to acknowledge the culture of white privilege, white entitlement, and racial superiority and exercise appropriate social responsibility. The authors cite the antiracist scholar Tim Wise's argument that the "ability of whites to deny nonwhite reality …is as strong as any other evidence of just how pervasive white privilege is in this society" (2005, 59). Overall, the text supports the belief that institutions are not neutral racial and gender spaces and states that we need to be color conscious and acknowledge biases that exist (Gutierrez y Muhs et al., 2012).

Byrd (2018) also supports the belief that student evaluations are biased, which is more apparent concerning female faculty and faculty of color. The author also notes that course concepts and students' perceptions of faculty often influence these evaluations (Byrd, 2018). The University of Dayton (n.d.) also discusses factors that can bias SET results and cite studies that revealed that the gender and ethnicity of students completing the SET may affect ratings and state that non-white faculty often receive lower ratings than white faculty.

Spooren et al. (2015) conducted a review of literature on SET in higher education and address the validity of SET. The authors acknowledge the fact that faculty have questioned the validity and reliability of SET. They cite a number of research studies that evaluated the validity of student ratings and the relationship between student biases and student evaluations. Overall, the research found that SETs yield useful information for faculty and administration. However, it is evident that there is wide variation among SET instruments. While several validated measures are available, many are developed without theory of effective knowledge as their foundation and lack content validity (Spooren et al., 2015). Some instruments may also lack face validity and not reflect students' perspectives of effective teaching. When considering convergent validity, the authors state that students' grades are often correlated with SET scores.

The research also reveals that expert ratings of faculty's teaching performance and self-ratings are correlated with SET scores, which provide support for convergent validity of SET.

Overall, it is apparent that there is a need to confirm the validity and reliability of SET used to evaluate faculty performance. It is also vital that potential biases impacting SET are considered. Acknowledging potential biases associated with SET and course evaluations is the first step in addressing this problem, but additional research and discussion concerning this issue is necessary.

Since in many institutions of higher learning, course evaluations or student ratings of faculty are used in the decision-making process for promotion and tenure, the following recommendations may be useful for administrators and those involved in the EDO and RTP process:

1. Bias in course evaluations may decrease with increased representation of women and minority groups in academia (Fan et al., 2019).
2. "…faculty members within departments and colleges—not administrators—should develop instruments and determine practices (peer review, classroom visits, teaching portfolios) that reflect the kinds of courses being taught, the levels of the students in the courses, and the styles of teaching being promoted. University-wide or college-wide evaluation forms that disregard this variety should be avoided; they generate meaningless numerical comparisons that invite misuse." (Vasey et al., 2016)
3. "…student evaluations should be completed in class. The move to online evaluations completed outside of class appears to compromise whatever reliability one could hope to claim for student evaluations." (Vasey et al., 2016)
4. "It may be possible to improve evaluations by trying to account for, or eliminate, bias — institutions could, for example, change the questions or discount the numbers to account for bias. But the better approach is to look at alternative means of assessing faculty performance." (Falkoff, 2018)

References

Aruguete, M. S., Slater, J., and Mwaikinda, S. R. (2017). The Effects of Professors' Race and Clothing Style on Student Evaluations. The Journal of Negro Education, 86(4), 494-502. http://www.jstor.org/stable/10.7709/jnegroeducation.86.4.0494

Byrd, W. C. (2018). From potential bias to action. https://www.insidehighered.com/advice/2018/05/04/how-avoid-bias-faculty-evaluations-opinion

Falkoff, M. (2018). Why We Must Stop Relying on Student Ratings of Teaching. The Chronicle of Higher Education. https://www.chronicle.com/article/why-we-must-stop-relying-on-student-ratings-of-teaching/

Fan Y, Shepherd LJ, Slavich E, Waters D, Stone M, Abel R, et al. (2019) Gender and cultural bias in student evaluations: Why representation matters. PLoS ONE 14(2): e0209749. https://doi.org/10.1371/journal.pone.0209749\

Gutierrez y Muhs, G., Niemann, Y. F., Gonzales, C. G., Harris, A. P., & Gonzales, C. G. (2012). *Presumed incompetent: The intersections of race and class for women in academia.* Utah State University Press. https://ebookcentral.proquest.com/lib/utc-ebooks/detail.action?docID=3442895&pq-origsite=primo

MacNell, L., Driscoll, A. & Hunt, A.N. (2015) What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. Innov High Educ 40, 291–303. https://doi.org/10.1007/s10755-014-9313-4

Mitchell, K., & Martin, J. (2018). Gender Bias in Student Evaluations. PS: Political Science& Politics, 51(3), 648-652. doi:10.1017/S104909651800001X

Reid, L. (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. doi: 10.1037/a0019865

Spooren, P., Brockx, B. S., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*(4), 598-642. doi:10.3102/0034654313496870

University of Dayton. (n.d.). Factors that can bias SET results. https://udayton.edu/ltc/set/faculty/bias.php

Vasey, C, & Carroll, L. (2016). How Do We Evaluate Teaching? American Association of University Professors 102(3). https://www.aaup.org/article/how-do-we-evaluate-teaching#.X5iBqUJKj0v

Wise, T. (2005). *White like me: Reflections on race from a privileged son*. Soft Skull Press.