How to interpret and use MEQ feedback

Student evaluations of modules are inherently biased. Designed with the aim to measure teaching quality, student ratings are also influenced by a range of other factors (e.g., Basow & Martin, 2012; Fan et al., 2019). One class of factors influencing student evaluations concerns instructors' characteristics, including but not limited to protected characteristics, such as gender, age, disability, ethnicity, accent, sexuality, social class, religion, and perceived ideology. Of these, gender bias is the most intensively researched and widely documented (MacNell, Driscoll & Hunt, 2015; Mitchell & Martin, 2018). Given the large proportion of instructors affected by gender bias, attempts have been made to quantify and correct for it statistically. However, such attempts have revealed the complex interaction of factors determining the extent to which gender bias will affect students' ratings and satisfaction with a module. Among these are raters' own gender, the subject taught, and the cultural context (Boring, Ottoboni, & Stark, 2016; Fan et al., 2019). Research has also revealed that (gender) bias can influence not only ratings on more impression-based dimensions, such as how engaging an instructor is, but also ratings on objective questions, such as timeliness of feedback (Boring et al., 2016).

Converging findings also emerge from research on bias in student evaluations based on ethnicity, age, accent, disability, sexuality, religion, etc. (Basow, Codos & Martin, 2013; Fan et al.,2019; Subtirelu, 2015), indicating the wide range of characteristics that can result in biased ratings. It is justified to assume that any social dimension along which prejudice and stereotyping can exist is likely to lead to evaluation bias, the magnitude of which is hard to predict and correct for. Therefore, student evaluations of modules should be interpreted and used with caution.

Below are some evidence-based guidelines on how to use student ratings (see Berk, 2018; Linse, 2017, for further detail).

Quantitative ratings

Take into consideration response rate and ratings distribution. A low response rate poses a question about the extent to which the ratings received are representative of the prevailing views within a class. The distribution of ratings also matters as the mean score can be significantly affected by a few extreme scores (i.e., outliers). Response rate and distribution should therefore be taken into account when interpreting student ratings.

Bias alone cannot account for excellent or poor ratings. Although the influence of bias cannot be accurately quantified, it is extremely unlikely that very high or very low student ratings are due to bias alone. Excellent ratings should therefore be taken as indicative of good quality teaching. Ratings with a mean below the midpoint of the scale (when the response rate and distribution are taken into account) indicate the presence of issues that have occurred in the teaching process and should be addressed.

Avoid comparisons among modules and instructors. Because the effect size of bias varies as a function of student cohort, subject taught, and instructor's characteristics, it is unclear to what extent it affects the ratings received by a given module or instructor. Members of faculty who are most likely to be affected are those who do not fit the stereotype of a university lecturer. It is therefore not justifiable to compare or rank-order modules/instructors for any diagnostic purposes, including promotion decisions.

Examine patterns of student ratings over several years and across modules to inform personnel decisions. Some differences in mean (average) ratings are common and may not be necessarily meaningful. Unusually low student ratings on a given module may also occur due to contextual factors (e.g., a problem with an assignment, industrial action) without being indicative of the instructor's overall teaching effectiveness. Taking into consideration the outcomes of teaching evaluations obtained over several years can therefore lead to a more accurate appraisal.

Qualitative comments

Focus on frequently occurring comments

Research indicates that written comments are highly correlated with student ratings. In forming a view on teaching quality, attention should be directed to frequently occurring comments, as they are representative of the students' perceptions. Isolated comments, whether positive or negative, should not be given undue weight as they may lead to misrepresentation of the collective student experience of the module.

Using student comments for module improvement. A good way to ensure that summaries of comments represent students' views is to sort student comments into groups based on similarity and label the group with a theme, then rank the themes based on the frequency of comments in each. Note that many students include multiple topics in a single sentence so those should be broken down into topical fragments and each sorted separately. Faculty members should focus improvement efforts on the first two to three themes, not the most negative comment.

Other indicators of teaching quality

Use student ratings as one source among others in appraising teaching quality. Student ratings and comments should be used in combination with other sources of teaching quality, such as internal and external reviewers' evaluation of a module, peer or mentor observations, teaching portfolios, students' learning as indicated by tests and assessments, as well as students' performance on subsequent modules that rely on knowledge obtained in the module under evaluation. These additional sources provide an important point of comparison to students' perspective. Data collection for each of these additional data sources should be systematic rather than informal.

References

Basow, S. A., & Martin, J. L. (2012). Bias in student evaluations. In M. E. Kite (Ed.), Effective evaluation of teaching: A guide for faculty and administrators (pp. 40–49). E-book retrieved from http://teachpsych. org/ebooks/evals2012/index.php

Basow, S., Codos, S., & Martin, J. (2013). The effects of professors' race and gender on student evaluations and performance. *College Student Journal*, *47*(2), 352-363.

Berk, R. A. (2018). Start spreading the news: Use multiple sources of evidence to evaluate teaching. *The Journal of Faculty Development*, *32*(1), 73-81.

Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*.

Fan, Y., Shepherd, L. J., Slavich, E., Waters, D., Stone, M., 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

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.

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40*(4), 291-303.

Mitchell, K. M., & Martin, J. (2018). Gender Bias in Student Evaluations. *PS: Political Science & Politics*, 1-5.

Subtirelu, N. C. (2015). "She does have an accent but…": Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors. com. *Language in Society*, *44*(1), 35-62.