

Frequently Quoted Bias Annotated Bibliography

Anderson, K. J., & Smith, G. (2005). Students' Preconceptions of Professors: Benefits and Barriers According to Ethnicity and Gender. *Hispanic Journal of Behavioral Sciences*, 27(2), 184–201.
<https://doi.org/10.1177/0739986304273707>

Abstract: The present study examined the influence of professor and student characteristics on students' preconceptions of college professors. Course syllabi for a politically charged social science course were constructed with versions varying by teaching style, professor gender, and professor ethnicity. A total of 633 (44% Latino; 34% African American; 22% Anglo) undergraduates rated the course and the instructor on professor warmth, professor capability, and political bias. Among several findings associated with professor ethnicity and teaching style, Latina professors were viewed as more warm when they had a lenient teaching style and less warm when they had a strict teaching style when compared with Anglo women professors with respective styles. Anglo men students perceived professors as more politically biased than did other students. Results are discussed in the context of aversive racism and a double standard of evaluation for Latino professors.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27–41.
<https://doi.org/10.1016/j.jpubeco.2016.11.006>

Abstract: This article uses data from a French university to analyze gender biases in student evaluations of teaching (SETs). The results of fixed effects and generalized ordered logit regression analyses show that male students express a bias in favor of male professors. Also, the different teaching dimensions that students value in male and female professors tend to match gender stereotypes. Men are perceived by both male and female students as being more knowledgeable and having stronger class leadership skills (which are stereotypically associated with males), despite the fact that students appear to learn as much from women as from men.

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.
<https://doi.org/10.1016/j.stueduc.2016.12.004>

Abstract: This article is about the accurate interpretation of student ratings data and the appropriate use of that data to evaluate faculty. Its aim is to make recommendations for use and interpretation based on more than 80 years of student ratings research. As more colleges and universities use student ratings data to guide personnel decisions, it is critical that administrators and faculty evaluators have access to research-based information about their use and interpretation. The article begins with an overview of common views and misconceptions about student ratings, followed by clarification of what student ratings are and are not. Next are two sections that provide advice for two audiences—administrators and faculty evaluators—to help them accurately, responsibly, and appropriately use and interpret student ratings data. A list of administrator questions is followed by a list of advice for faculty responsible for evaluating other faculty members' records.

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. <https://doi.org/10.1007/s10755-014-9313-4>

Abstract: Student ratings of teaching play a significant role in career outcomes for higher education instructors. Although instructor gender has been shown to play an important role in influencing student ratings, the extent and nature of that role remains contested. While difficult to separate gender from teaching practices in person, it is possible to disguise an instructor's gender identity online. In our 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. Given the vital role that student ratings play in academic career trajectories, this finding warrants considerable attention.

Mitchell, K. M. W., & Martin, J. (2018). Gender Bias in Student Evaluations. *PS - Political Science and Politics*, 51(3), 648–652. <https://doi.org/10.1017/S104909651800001X>

Abstract: Many universities use student evaluations of teachers (SETs) as part of consideration for tenure, compensation, and other employment decisions. However, in doing so, they may be engaging in discriminatory practices against female academics. This study further explores the relationship between gender and SETs described by MacNell, Driscoll, and Hunt (2015) by using both content analysis in student-evaluation comments and quantitative analysis of students' ordinal scoring of their instructors. The authors show 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. Findings suggest that the relationship between gender and teaching evaluations may indicate that the use of evaluations in employment decisions is discriminatory against women.

Reid, L. 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. <https://doi.org/10.1037/a0019865>

Abstract: The present study examined whether student evaluations of college teaching (SETs) reflected a bias predicated on the perceived race and gender of the instructor. Using anonymous, peer-generated evaluations of teaching obtained from RateMyProfessors.com, the present study examined SETs from 3,079 White; 142 Black; 238 Asian; 130 Latino; and 128 Other race faculty at the 25 highest ranked liberal arts colleges. Results 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. A two-stage cluster analysis demonstrated that the very best instructors were likely to be White, whereas the very worst were more likely to be Black or Asian. Few effects of gender were observed, but several interactions emerged showing that Black male faculty were rated more negatively than other faculty. The results of the present study are consistent with the negative racial stereotypes of racial minorities and have implications for the tenure and promotion of racial minority faculty.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). *On the Validity of Student Evaluation of Teaching: The State of the Art*. *Review of Educational Research* (Vol. 83). <https://doi.org/10.3102/0034654313496870>

Abstract: This article provides an extensive overview of the recent literature on student evaluation of teaching (SET) in higher education. The review is based on the SET meta-validation model, drawing upon research reports published in peer-reviewed journals since 2000. Through the lens of validity, we consider both the more traditional research themes in the field of SET (i.e., the dimensionality debate, the 'bias' question, and questionnaire design) and some recent trends in SET research, such as online SET and bias investigations into additional teacher personal characteristics. The review provides a clear idea of the state of the art with regard to research on SET, thus allowing researchers to formulate suggestions for future research. It is argued that SET remains a current yet delicate topic in higher education, as well as in education research. Many stakeholders are not convinced of the usefulness and validity of SET for both formative and summative purposes. Research on SET has thus far failed to provide clear answers to several critical questions concerning the validity of SET.

Stark, P., Ottoboni, K., & Boring, A. (2016). Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness. *ScienceOpen Research*, 1–11. <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1>

Abstract: Student evaluations of teaching (SET) are widely used in academic personnel decisions as a measure of teaching effectiveness. We show: • SET are biased against female instructors by an amount that is large and statistically significant • the bias affects how students rate even putatively objective aspects of teaching, such as how promptly assignments are graded • the bias varies by discipline and by student gender, among other things • it is not possible to adjust for the bias, because it depends on so many factors • SET are more sensitive to students' gender bias and grade expectations than they are to teaching effectiveness • gender biases can be large enough to cause more effective instructors to get lower SET than less effective instructors. These findings are based on nonparametric statistical tests applied to two datasets: 23,001 SET of 379 instructors by 4,423 students in six mandatory first-year courses in a five-year natural experiment at a French university, and 43 SET for four sections of an online course in a randomized, controlled, blind experiment at a US university.

Stark, P., & Freishtat, R. (2014). An Evaluation of Course Evaluations. *ScienceOpen Research*, (September), 1–26. <https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1>

Abstract: Student ratings of teaching have been used, studied, and debated for almost a century. This article examines student ratings of teaching from a statistical perspective. The common practice of relying on averages of student teaching evaluation scores as the primary measure of teaching effectiveness for promotion and tenure decisions should be abandoned for substantive and statistical reasons: There is strong evidence that student responses to questions of “effectiveness” do not measure teaching effectiveness. Response rates and response variability matter. And comparing averages of categorical responses, even if the categories are represented by numbers, makes little sense. Student ratings of teaching are valuable when they ask the right questions, report response rates and score distributions, and are balanced by a variety of other sources and methods to evaluate teaching.

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*. <https://doi.org/10.1016/j.stueduc.2016.08.007>

Abstract: Student evaluation of teaching (SET) ratings are used to evaluate faculty's teaching effectiveness based on a widespread belief that students learn more from highly rated professors. The key evidence cited in support of this belief are meta-analyses of multisection studies showing small-to-moderate correlations between SET ratings and student achievement (e.g., Cohen, 1980, 1981; Feldman, 1989). We re-analyzed previously published meta-analyses of the multisection studies and found that their findings were an artifact of small sample sized studies and publication bias. Whereas the small sample sized studies showed large and moderate correlation, the large sample sized studies showed no or only minimal correlation between SET ratings and learning. Our up-to-date meta-analysis of all multisection studies revealed no significant correlations between the SET ratings and learning. These findings suggest that institutions focused on student learning and career success may want to abandon SET ratings as a measure of faculty's teaching effectiveness.