

Student Evaluations of Teaching (SET) Summary and Background
for the AAPA Executive Committee
as AAPA Considers Recommendations on Use and Best Practice
April, 2019

President Aiello asked the AAPA Committee on Diversity to provide background and recommendations to the AAPA Executive Committee on current research concerning Student Evaluations of Teaching (SET), their use, effectiveness, concerns and best practice. The undersigned committee comprises COD members (from a variety of subcommittees) as well as other AAPA members with relevant expertise and interest. We provide an overall assessment of SET, a synopsis of relevant literature on the topic and recommendations for SET use.

Overall:

- SET are biased against women faculty and faculty of color (as has been clearly demonstrated in multiple studies)
- Gendered expectations relating to professors' accommodation of student requests factors into lower SET scores
- Administrators are complicit in that they have often sought to use summary measures of teaching effectiveness, which are not statistically or otherwise justified
- Negative SET factor into hiring/promotion/tenure/merit review, although they are not clear measures of teaching effectiveness (they are measures of student opinion)
- Multiple measures of teaching, including peer review, are critical for obtaining a well-rounded and complete picture of teaching effectiveness
- Evaluations should center on the course and learning rather than the instructor *per se*

Research demonstrates that student evaluations of teaching (SET) are consistently biased against women faculty and faculty of color, and students bring differing expectations of professors based on gender, race, and other factors (Boring et al 2016, El-Alayli et al 2018, Pittman 2010, Reid 2010). These differing expectations of professors can lead to lower SET scores. Boring et al (2016) note that students' gendered expectations impact their evaluation of both objective factors, such as the promptness with which faculty return graded work, and subjective and largely emotional factors, such as the faculty member's enthusiasm and caring. El-Alayli et al (2018) also demonstrate that students expect women faculty to provide more emotional support and nurturing than their male counterparts. Although SET were originally created as a method of shaping the quality of teaching, they are now often used as the main method of evaluating teaching effectiveness (Hornstein 2017). Further, many interpreters of the data are untrained in how to understand SET results (Hornstein 2017). Research also demonstrates that aside from biases, SET do not measure teaching effectiveness. In one study, teaching effectiveness was measured by the researchers and compared with SET results. In this case, teaching effectiveness was actually negatively correlated with SET results (Braga et al 2014). In addition, factors such as weather and presence of cookies are correlated with SET scores (Braga et al 2014, Hessler et al 2018). In biological anthropology, the percentage of white women and racial minorities decreases at higher ranks; this pattern mirrors those seen in other scientific disciplines (Antón et al, 2018; Sheltzer and Smith 2014; Turner et al, 2018). Further, women are more likely to be hired into teaching-focused positions, where SET results may be weighed more heavily in hiring, contract renewal, promotion, and tenure (Sheltzer and Smith, 2014). Thus, the biases underlying SET may contribute to the perpetuation of racial and gender inequities in biological anthropology and other sciences.

With this in mind, **we advocate for limited use of SET in faculty hiring, contract renewal, promotion, tenure, and merit reviews.** Other methods of assessment, such as classroom observation, teaching portfolios, and peer evaluation of teaching materials are available and can provide more useful and less biased data (Hornstein 2017). Several universities in the United States have recently shifted toward new models of faculty evaluation that apply a more holistic approach and prioritize these alternate assessment tools (Flaherty 2018). In addition, if SET continue to be utilized, we encourage critical reflection on the questions used and training in the interpretation of the resulting data. SET questions should be carefully worded to avoid topics outside the students' experience and expertise, such as asking students whether the faculty member is "knowledgeable" about the subject, and to avoid misleading or vague judgments, such as asking students to evaluate a faculty member's "overall teaching effectiveness" (Hornstein 2017, Stark and Freishtat 2014). The data compiled via SET administration should also be carefully presented and interpreted. Score distributions should be used, rather than means; when response rates are low, data should not be extrapolated to the whole class; and courses of different types, levels, sizes, and disciplines should not be compared (Stark and Freishtat 2014).

References

- Antón, S.C. Malhi, R.S. & Fuentes, A. (2018). Race and diversity in U.S. Biological Anthropology: A decade of AAPA initiatives. *American Journal of Physical Anthropology* 165: 158–180.
- Boring, A. Ottoboni, K. & Stark, P.B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpenResearch*. DOI:10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1
- Braga, M. Paccagnella, M. & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review* 41:71-88.
- El-Alayli, A. Hansen-Brown, A.A. & Ceynar, M. (2018). Dancing backwards in high heels: female professors experience more work demands and special favor requests, particularly from academically entitled students. *Sex Roles* 79:136–150.
<https://doi.org/10.1007/s11199-017-0872-6>
- Flaherty, C. (2018). Teaching eval shake-up. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2018/05/22/most-institutions-say-they-value-teaching-how-they-assess-it-tells-different-story?utm_source=Inside+Higher+Ed&utm_campaign=92984764da-DNU&utm_medium=email&utm_term=0_1fcbc04421-92984764da-233561669&mc_cid=92984764da&mc_eid=9c69010b52.
- Hessler, M. Pöpping, D.M. Hollstein, H. Ohlenburg, H. Arneman, P.H. Massoth, C. Seidel, L.M. Zarbock, A. & Wenk, M. (2018). Availability of cookies during an academic course session affects evaluation of teaching. *Medical Education* 52: 1064-1072. <https://doi.org/10.1111/medu.13627>
- Hornstein, H.A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education* 4: 1304016.
- Pittman, C.T. (2010). Race and gender oppression in the classroom: the experiences of women faculty of color with white male students. *Teaching Sociology* 38(3) 183–196. DOI: 10.1177/0092055X10370120
- 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: 137–152. DOI: 10.1037/a0019865
- Sheltzer, J.M. & Smith J.C. (2014). Elite male faculty in the life sciences employ fewer women. *PNAS* 111: 10107–10112.

Stark, P.B. & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpenResearch*. DOI: 10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

Turner, T.R. Bernstein, R.M. Andrea B. & Taylor, A.B. (2018). Participation, representation, and shared experiences of women scholars in biological anthropology. *American Journal of Physical Anthropology* 165: 126–157.

Contributors:

Laurie Kauffman (COD AACT)
Department of Biology
Oklahoma City University

Susan Antón (COD IDEAS)
Department of Anthropology
New York University

Summer Arrigo-Nelson (COD AACT)
Department of Biological and Environmental Science
California University of Pennsylvania

Robin Bernstein (COD WIN)
Department of Anthropology
University of Colorado

Michelle A. Rodrigues
Beckman Institute for Science and Technology
University of Illinois, Urbana-Champaign

Chris Robinson
Department of Biological Sciences
Bronx Community College, City University of New York

K. Elizabeth Soluri
Department of Anthropology
Cabrillo College