

White Paper on Use of Student Opinion Questionnaires to Evaluate Teaching Effectiveness at California State University Fullerton

*Researchers and Critical Educators (RACE)*¹

Introduction

Researchers and Critical Educators is a university-wide organization that includes 164 members, including pre-tenure, tenured, and non-tenure track faculty from all colleges within California State University, Fullerton (CSUF). We support faculty on campus and promote goals related to diversity and social justice.

The purpose of this paper is to challenge the overreliance on student evaluations within the retention, tenure and promotion (RTP) process at CSUF. Such evaluations, called student evaluations of teaching (SETs) or student opinion questionnaires (SOQs), have been widely criticized as both invalid measures of student learning and sources of significant bias (Arbuckle & Williams, 2003; Riniolo, Johnson, Sherman & Misso, 2006; Williams, 2007). As a collective, we are troubled by research indicating that SETs are heavily influenced by racial and gender bias and thus are improper tools within the RTP process at CSUF.

In this paper, we briefly summarize the ways SOQs are used at CSUF, describe the research regarding bias in SETs nationwide, and close with recommendations for action at the departmental, college and university levels.

Background on Student Evaluation of Teaching at CSUF

Student evaluations of teaching (SETs), called Student Opinion Questionnaires (SOQs) at California State University, Fullerton, are end-of-semester surveys used to evaluate the teaching effectiveness of individual faculty members. The CSU Collective Bargaining Agreement (CBA) states that periodic evaluation procedures “shall, for tenure-track faculty unit employees who teach, include, but not be limited to, student evaluations of teaching performance” (CBA, p. 52). Individual departments determine the specific questions included on SOQs, and the degree of emphasis placed on SOQs relative to other measures of teaching effectiveness. An informal survey of RACE constituents underscores the wide disparity in the types of questions included on SOQs, the weights placed upon them, and the implementation processes campus-wide.

¹For inquiries, please contact RACE E-Board

Moreover, university-level research regarding SOQs underscore the gross disparities in response rates university-wide, underscoring systemic issues related to the ways SETs are administered and used at CSUF.

In this paper, however, we wish to focus on one specific concern with the use of SOQs at CSUF: an overreliance on SOQs as a criterion for teaching effectiveness/performance creates a condition of bias which affects the evaluation of faculty during RTP processes. This is in conflict with CSUF, CSU and state equity principles (see CSUF strategic plan on [diversity, equity and inclusion](#); [CSU HR statement on hiring](#); and [the CA Dept of Ed statement on gender equity](#)), and may also have legal implications (see [arbitration at Ryerson University](#)).

Racial and Gender Bias in SETs

There is a plethora of evidence that SETs are biased against women (Basow & Silberg, 1987; Boring, Ottoboni & Stark, 2016; MacNell, Driscoll & Hunt, 2015; Mitchell & Martin, 2018) and may be in part responsible for the under-representation of women in senior academic roles (particularly in male-dominated fields) (Burnell, Cojuharenco, & Murad, 2018). For example, MacNell et al., (2015) examined student ratings of assistant instructors with perceived gender difference within an online course. The study found that “students rated male identities significantly higher than female identities, regardless of the instructors’ actual gender” (MacNell et al., 2015, p. 291).

In addition, evidence shows that SETs are biased against faculty of color, and in particular, African American, Latina/o and Asian faculty (Anderson & Smith, 2005; Littleford & Jones, 2017; Reid, 2010; Smith & Hawkins, 2011) and faculty whose first language isn’t English (Fan et.al.,2019). African American and Asian faculty were more likely to receive negative evaluations in terms of “overall quality, helpfulness and clarity” despite having similar abilities and credentials as their White counterparts (Reid, 2010, p. 137). Further, African American faculty scored lower mean scores on global SET items (i.e., overall teaching ability and overall value of the course) than other measured faculty racial/ethnic groups, including White, Asians and Latinos (Smith & Hawkins, 2011). This finding was counterintuitive because African Americans had high scores on multidimensional items (i.e., the instructor was clear about basic principles and I felt free to voice my opinion in this course) (Smith & Hawkins, 2011, pp. 154-156) indicating a clear bias against African American faculty. Thus, CSUF’s reliance upon SETs as a significant indicator in RTP processes sets up an inequitable system in which student bias

¹For inquiries, please contact RACE E-Board

has a direct impact upon the university's evaluation of faculty members, with negative implications for women and people of color. Moreover, these institutional policies undermine CSUF's efforts to recruit and retain highly qualified and diverse faculty, as detailed in Strategic Plan Goal #3 (<http://planning.fullerton.edu/2018-2023-plan/goal3.aspx>).

Research suggests the impact of bias in SETs is especially significant for faculty teaching coursework related to diversity, including topics like race, language, (dis)ability, gender, social justice, community engagement and its intersections, topics mostly covered by faculty of color. Littleford & Jones (2017), for example, found that students perceive faculty of color discussing these "controversial issues" as biased, resulting in lower SETs. Together with the lower evaluations of faculty of color in general, bias is compounded and most disproportionately discriminates against women of color.

Given the evidence of gender and racial bias, we maintain that SETs should not be used to evaluate faculty for retention, tenure or promotion. Beyond the obvious implications on job security, professional advancement and income inequality, the well-established bias against women and faculty of color renders SETs incompatible with the equity principles of CSUF, CSU and the state of California.

These equity concerns go beyond mere academic ethics and notions of fairness. There has been a recent legal arbitration decision at Ryerson University that SOQs can no longer be used to evaluate teaching effectiveness for RTP purposes (Doerer, 2019). This likely only represents the beginning as larger lawsuits appear inevitable. As one prominent scholar studying SETs put it: "It's just a question of time before there are class-action lawsuits against universities or even whole state-university systems on behalf of women or other minorities, alleging disparate impact" (Philip Stark quoted by Doerer, 2019 in the Chronicle of Higher Education).

We therefore recommend that CSUF, in accordance with its equity principles, stop using SETs for decisions regarding retention, tenure and promotion.

Recommendations

Given the established bias and questionable legality of SETs, we as Members of RACE provided a list of Recommendations to help the university move forward in this complex work of judging effective teaching.

- Stop using SOQs as a measure of teaching effectiveness. Instead, explicitly position SOQs as an analysis of students' *opinions* of teaching.

¹For inquiries, please contact RACE E-Board

- Use other, validated, measures of evaluating teacher effectiveness including but not limited to observations and examinations of teaching materials (classroom activities, assessments, etc.). These measures should be weighted *more heavily* than SOQs in RTP processes.
- The onus should be on universities that rely on SOQs for employment decisions to provide convincing affirmative evidence that such reliance does not have disparate impact on women, underrepresented minorities, or other protected groups. Because the bias varies by course and institution, affirmative evidence needs to be specific to a given course in a given department in a given university. Absent such specific evidence, SET should not be used for personnel decisions (Boring, Ottoboni & Stark, 2016, p. 11).
- Explore and designate funds to support the development and analysis of alternatives to SOQs, such as teaching portfolios, teaching demonstrations, videos, performance based measures, etc.
- Work with CFA to modify CBA articles on student evaluations of teaching performance and align UPS with any CBA changes.
- DPC committees should observe or arrange observations of classroom/online teaching of tenure track/non-tenure track faculty at least 2x per course per academic year to collect additional data to properly evaluate RTP portfolios.

Conclusions

SOQs are not an acceptable criterion for RTP decisions because of their established gender and racial bias. This is in violation of CSUF and CSU equity principles and likely presents grounds for legal challenges in the State of California. Alternative modes and criteria for evaluations should be explored and implemented. We must engage in the complex work of agreeing on what Teaching Effectiveness means and the various ways that faculty can demonstrate being on the right track towards reaching this standard. Doing this in a manner that honors the diversity, skills and strengths of diverse faculty will be in harmony and support of CSUF equity principles and its overall mission.

References

- 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. doi:10.1177/0739986304273707
- Arbuckle J., & Williams B. D. (2003). Students' perceptions of expressiveness: age and gender effects on teacher evaluations. *Sex Roles*, 49: 507–516.
Doi.org/10.1023/A:1025832707002
- 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.
- Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology*, 79(3), 308–314.
Doi:10.1037/0022-0663.79.3.308
- Boring, A., Ottoboni, K., & Stark, P. (2016). Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness. *ScienceOpen Research*. 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. Doi:<https://doi-org.lib-proxy.fullerton.edu/10.1016/j.econedurev.2014.04.002>
- Burnell, N., Cojuharenco, I., & Murad, Z. (2018). *Now you see it not you don't: The effect of teaching style and seniority on gender bias in teaching evaluations*. (Discussion Paper Series No. DP No.09/2018). Surrey, UK: University of Surrey. Retrieved from <http://www.deliveringbetter.com/uploads/2/6/8/0/26806980/dp-no-9-teaching-evaluations.pdf>
- Collective Bargaining Agreement between the Board of Trustees of the California State University and the California Faculty Association, Unit 3 Faculty, November 12, 2014 – June 30, 2017 (Extended to June 30, 2020).
- Collom, E., & Calucag, N. (2019). Student Opinion Questionnaires: Overview, Processes, and Response Rates [Presentation]. Retrieved from <http://www.fullerton.edu/far/soq/>.
- Doerer, K. (2019). Colleges are getting smarter about student evaluations: Here's how. *The Chronicle of Higher Education*, , Apr. 15, 2019.
- Fan, Y., Shepherd, L. J., Slavich, E., Waters, D., Shone, M., Abel, R., & Johnson, E. L. (2019).

¹For inquiries, please contact RACE E-Board

- Gender and cultural bias in student evaluations: Why representation matters. *Plos One*, 14(2), 1-16. doi:<https://doi.org/10.1371/journal.pone.0209749>
- Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4, 1-8.
- Littleford, L. N., & Jones, J. A. (2017). Framing and source effects on white college students' reactions to racial inequity information. *Cultural Diversity and Ethnic Minority Psychology*, 23(1), 143-153.
- MacNell, L., Driscoll, A., & Hunt, A. N. (2014). What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. *Innovative Higher Education*, 40(4), 291–303.
Doi:10.1007/s10755-014-9313-4
- Mitchell, K. M. W., & Martin, J. (2018). Gender Bias in Student Evaluations. *PS: Political Science & Politics*, 51(03), 648–652. Doi:10.1017/s104909651800001x
- 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.
Doi:10.1037/a0019865
- Riniolo, T. C., Johnson, K. C., Sherman, T. R., & Misso, J. A. (2006). Hot or Not: Do Professors Perceived as Physically Attractive Receive Higher Student Evaluations? *The Journal of General Psychology*, 133(1), 19–35. Doi:10.3200/genp.133.1.19-35
- 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.
- Stark, P., & Freishtat, R. (2014). An evaluation of course evaluations. *Science Open*, doi: 10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1
- Wigington, H., Tollefson, N., & Rodriguez, E. (1989). Students' ratings of instructors revisited: Interactions among class and instructor variables. *Research in Higher Education*, 30(3), 331-344. doi:10.1007/BF00992608
- Williams, D. A. (2007). Examining the Relation between Race and Student Evaluations of Faculty Members: A Literature Review. *Profession*, 2007(1), 168–173.
Doi:10.1632/prof.2007.2007.1.168

Appendix

RACE E-Board Members 2019-2020

Alison Dover - Co-president

Benikia Kressler - Co-president

Maritza Lozano - Secretary

Nick Henning – Treasurer

Aitana Guia – Member at Large

Ricardo Valencia – Writing Group Chair

Estela Zarate – Public Relations Chair

September 2019

¹For inquiries, please contact RACE E-Board