

C-FSC Educational Policies and Faculty/Student Relations Committee

Date: November 05, 2019 DRAFT

Members: Scott Illingworth (Chair), Alison Aldrich, John Gershman, Noelle Molé Liston, Iskender Sahin

REPORT

John Gershman shared the included policy brief produced by a colleague's student titled, "Student Evaluations of Teaching: An Enemy in Disguise" that relates to recent conversations on the matter.

The university is engaged in looking at Course evaluations from two angles:

1. An ad-hoc University wide committee on redesigning student evaluations has been created with a focus on people with expertise on metrics and survey design. Ethan Youngerman is the C-FSC representative on that committee. That committee's focus is, as we understand it, to focus solely on the design of student evaluations.
2. The Senate's Academic Affairs Committee is establishing a sub-committee that will explore best practices in the evaluation of teaching (portfolios, peer observation, etc) to make a recommendation regarding the best way(s) to evaluate teaching with respect to reappointment, promotion, and tenure. That sub-committee has yet to meet.

We will also seek the findings of the internal study related to bias in course evaluations and will share anything we are given with the council.

Members of the Ed Policy Committee are scheduled to meet with reps from SSC on November 11, 2019 and we can report about that conversation verbally at the meeting.

We are working to schedule a meeting with NJ Knoll Finn and her team. Some of the items for conversation include:

- Update on the pilot version of the spring admits. Sharing additional questions.
- Rollout of Academic Engagement and Gradebook data in NYU Classes
- Questions regarding data sharing agreements with Starfish (the back end software for NYUConnect)

Other council members should advise if there are additional matters they wish is to raise in that meeting.

Respectfully submitted,

Scott Illingworth

STUDENT EVALUATIONS OF TEACHING: AN ENEMY IN DISGUISE

The Limitations of SETs in Assessing Instructor Performance

Background

One of the major factors contributing to the disparities in tenure, promotion, and pay decisions among teachers is the over-reliance on student evaluations of teaching (SETs) as measures of an instructor's effectiveness [31, 46]. Research has shown that SETs are flawed tools of measurement as they are highly subjective and susceptible to bias [39, 41].

What are Student Evaluations of Teaching?

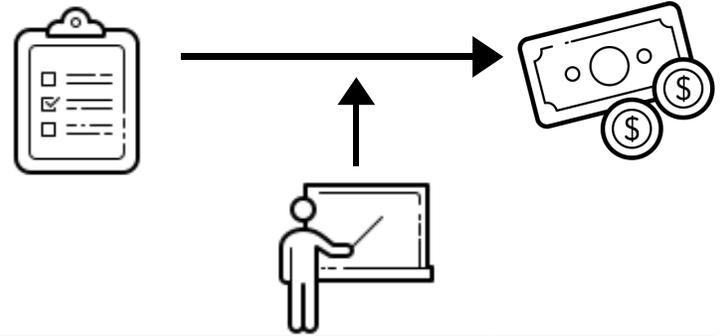
The feedback students provide at the end of a course. Usually presented in the forms of surveys or questionnaires, they are used to evaluate a student's learning and how beneficial they found the course.

What do SETs measure?

Teaching evaluations are poor indicators of teacher quality as the results are usually the outcomes of students' preferences. For example, classes with lighter workloads and higher grading distributions are more likely to get higher scores [22, 29, 37].

Evaluations are also higher when students feel engaged in the course, thus non-elective courses and quantitative courses score much lower [8, 10, 15, 18, 22, 28, 44, 45].

Additionally, disciplinary differences are present as natural sciences usually have the lowest scores while humanities courses have the highest [21, 22, 29, 39, 44, 45, 48]. Therefore, evaluations are based more on course-specific characteristics than they are on instructor quality and performance.



Bias in SETs have material consequences for those who are negatively affected, such as job security, pay increases and opportunities for promotion.

Are SETs Effective Tools of Measurement?

SETs are not practical measures of teacher effectiveness as they are primarily focused on reliability; they show whether different students give their instructor similar evaluations [41]. However, SETs are not valid measures of whether an instructor helped or hindered a student's learning. Controlled randomized experiments measuring how valuable SETs are at assessing instructor efficiency actually found that SET ratings are negatively associated with direct measures of effectiveness [41, 44].

What SETs do measure are the following:

- Students' perceptions of how well they will do in the course [22, 37, 41]
- How much they enjoyed the material [22, 41, 44]
- Their personal biases towards their instructor [31, 41]

These are not adequate measures of how well an instructor guided a student's comprehension of the course material.



Gender-Based Discrepancies in Evaluation Results



- Men are more likely to be perceived as more accurate in their teaching, competent, organized and professional compared to their female counterparts [1, 4, 5, 6, 11, 18, 27, 28, 34, 36, 37, 40, 42].
- In online teaching environments, students are more likely to offer lower evaluations if they perceive the instructor to be a woman [11, 27].
- In a study conducted using SETs and course evaluations, the results found that men who were native-English speakers were most likely of getting the highest scores on their evaluations in every subject area except Engineering [20].
- Gender of the student affects how they evaluate their instructors as male students rate their male instructors higher and their female instructors lower, while the opposite is true for female students [5, 6, 12, 13, 14, 20, 23, 28, 35].

Identity-Based Biases

- Faculty of color and faculty with accents get lower scores compared to their white counterparts. One study shows that Black and other non-white faculty received the lowest mean scores across various evaluation items of course quality, overall value and overall teaching ability [38]. Faculty with accents and Asian last names receive lower SET and Rate My Professor ratings as well [20, 43].
- Experimental evidence on biases towards LGBTQIA+ faculty found that students were more likely to rate gay and lesbian faculty who were strong lecturers lower than lecturers with an unspecified orientation, though they also rated gay and lesbian faculty who were weak lecturers more moderately [2, 19].
- The research on the possibility of sexuality and racial bias in SETs is limited because these populations are severely underrepresented in academia [3]. In order to improve this research, there needs to be a concerted effort to hire and support LGBTQIA+ faculty and faculty of color.

Role of Stereotypes in Evaluations

Women and men appear to be evaluated on their compliance, or lack thereof, to gender stereotypes [47]. Women are evaluated higher if they exhibit traits associated with femininity, such as warmth and compassion [7, 16, 25, 36]. Men, on the other hand, are judged on their perceived intellectual and teaching capabilities [6, 9, 10, 26]. In addition, students expect more subjective treatment from female instructors and react negatively when those expectations are not satisfied [17, 33]. Thus, whether or not an instructor confirms to their prescribed gender roles has a stronger effect on their evaluation scores than their actual performance.

Implications

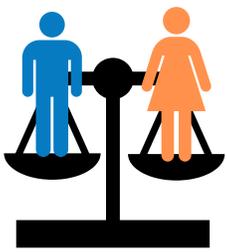
The reliance on SETs as accurate measures of teacher effectiveness has significant ramifications for who benefits and who does not [39, 41, 46]. As women and faculty of color are less likely to get high scores on evaluations they receive from students, they are less likely to reap benefits such as higher pay and promotional opportunities compared to their white, male counterparts. For example, in Political Science, disparities such as these result in fewer women being represented at each stage, from the graduate level to faculty and professional levels [3]. These discrepancies lead to tangible consequences that result in the lack of women and people of color in higher positions.

Recommendations for Improvement of Teaching Evaluations

Addressing Student Bias

Directly addressing the gender bias students carry is one way to decrease the discrepancies in evaluation results between male and female instructors [24, 32].

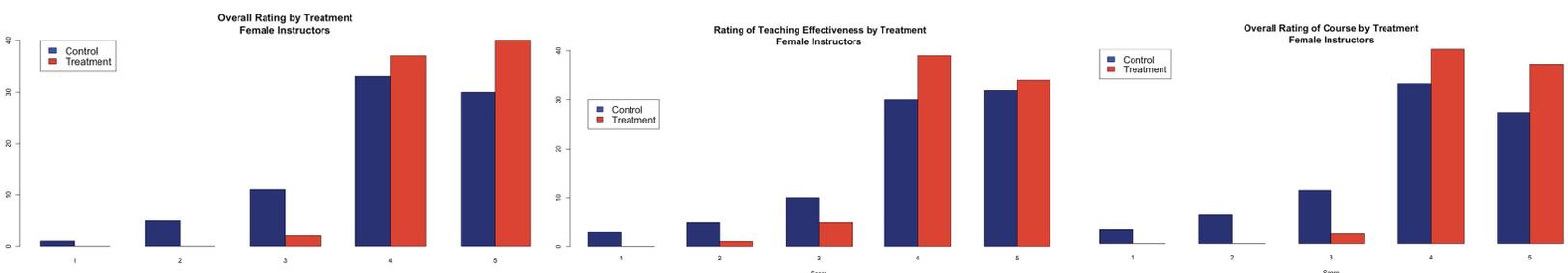
Research has been conducted to test whether or not informing students of the potential for gender biases has any effect on the evaluation results of teachers and the results are compelling. Evaluation scores for female instructors increased by as much as half a point on a five-point scale [32]. However, research on this topic is fairly new and limited in scope.



Reformatting Survey Questions

The design of the tools used to collect SETs have a significant affect on the presence of gender bias in the results. By making alterations, these tools can be mitigate the inequalities between male and female instructors. In a study published in the American Sociological Review, reducing the number of scale points from 10 to 6 helped shrink the gender gaps in teachers' performance evaluations [34]. In a separate study, eliminating global survey questions and including multidimensional items also limit the bias present in SETs [38].

Figure 2: SET Results by Experimental Condition



A study conducted at Iowa State University by Fan et al. found that students who were informed of potential bias rated their female instructors higher than those who did not receive such information [32].

Measuring Course Satisfaction

Students are in a good position to observe and evaluate aspects of the course outside of the instructor. Student-centered evaluations can be predictive of how effective the course is overall. SETs measure criteria like workload and course enjoyment [22, 37, 41, 44]. By formatting survey questions to include the pace of the class and whether or not the course will have any value in their future careers, this tool can be more useful and forecast what improvements the instructor can make.

Utilizing Peer Evaluations

Using faculty members to be a part of the evaluation process for instructors could present a more holistic approach to measuring teaching effectiveness (41). Unlike SETs, faculty can have a broader perspective on what is and is not effective in the classroom. However, the potential for racial and gender based bias is still present. Faculty who are observing should know of these potential biases and take steps to mitigate them, such as taking an implicit bias test or a multicultural competency course.

For more information and research on teaching evaluations, please visit <https://www.rebeccakreitzer.com/bias/>

Work Cited

1. Abel, Millicent H, and Andrea L Meltzer. 2007. "Student ratings of a male and female professors' lecture on sex discrimination in the workforce." *Sex Roles* 57 (3-4):173-180.
2. Anderson, Kristin J, and Melinda Kanner. 2011. "Inventing a Gay Agenda: Students' Perceptions of Lesbian and Gay Professors 1" *Journal of Applied Social Psychology* 41 (6):1538-1564.
3. APSA. 2011. *Political Science in the 21st Century*. edited by Report of the Task Force on Political Science in the 21st Century.
4. Arbuckle, Julianne, and Benne D Williams. 2003. "Students' perceptions of expressiveness: Age and gender effects on teacher evaluations." *Sex Roles* 49 (9-10):507-516.
5. Basow, Susan A. 1995. "Student evaluations of college professors: When gender matters." *Journal of Educational Psychology* 87 (4):656.
6. Basow, Susan A. 2000. "Best and worst professors: Gender patterns in students' choices." *Sex roles* 43 (5-6):407-417.
7. Bennett, Sheila K. 1982. "Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation." *Journal of Educational Psychology* 74 (2):170.
8. Benton, SL, and WE Cashin. 2012. *Student ratings of teaching: A summary of research and literature* (IDEA Paper no. 50). Manhattan, KS: The IDEA Center.
9. Bian, Lin, Sarah-Jane Leslie, and Andrei Cimpian. 2017. "Gender stereotypes about intellectual ability emerge early and influence children's interests." *Science* 355 (6323):389-391.
10. Boring, Anne. 2017. "Gender biases in student evaluations of teaching." *Journal of public economics* 145:27-41.
11. Boring, Anne, Kellie Ottoboni, and Philip Stark. 2016. "Student evaluations of teaching mostly do not measure teaching effectiveness." *ScienceOpen Research*.
12. Bray, James H, and George S. Howard. 1980. "Interaction of teacher and student sex and sex role orientations and student evaluations of college instruction." *Contemporary Educational Psychology* 5 (3):241-248.
13. Burns-Glover, Alyson L, and Dale J Veith. 1995. "Revisiting gender and teaching evaluations: Sex still makes a difference." *Journal of Social Behavior and Personality* 10 (4):69.
14. Centra, John A, and Noreen B Gaubatz. 2000. "Is there gender bias in student evaluations of teaching?" *The journal of higher education* 71 (1):17-33.
15. Chamberlin, Marilyn S, and JoAnn S Hickey. 2001. "Student evaluations of faculty performance: The role of gender expectations in differential evaluations." *Educational Research Quarterly* 25 (2):3.
16. Eagly, Alice H, and Steven J Karau. 2002. "Role congruity theory of prejudice toward female leaders." *Psychological review* 109 (3):573.
17. El-Alayli, Amani, Ashley A Hansen-Brown, and Michelle Ceynar. 2018. "Dancing backwards in high heels: Female professors experience more work demands and special favor requests, particularly from academically entitled students." *Sex Roles* 79 (3-4):136-150.
18. Elmore, Patricia B, and Karen A LaPointe. 1975. "Effect of teacher sex, student sex, and teacher warmth on the evaluation of college instructors." *Journal of Educational Psychology* 67 (3):368.
19. Ewing, Vanessa Lynn, Arthur A Stukas Jr, and Eugene P Sheehan. 2003. "Student prejudice against gay male and lesbian lecturers." *The Journal of Social Psychology* 143 (5):569-579.
20. Fan, C, LJ Shepherd, E Slavich, D Waters, M Stone, R Abel, and EL Johnston. 2019. "Gender and cultural bias in student evaluations: Why representation matters." *PLoS one* 14 (2):e0209749.
21. Franklin, Jennifer, and Michael Theell. 1995. "The relationship of disciplinary differences and the value of class preparation time to student ratings of teaching." *New directions for teaching and learning* 1995 (64):41-48.
22. Greenwald, Anthony G, and Gerald M Gillmore. 1997. "No pain, no gain? The importance of measuring course workload in student ratings of instruction." *Journal of Educational Psychology* 89 (4):743.
23. Kaschak, Elynn. 1981. "Another Look at Sex Bias in Students' evaluations of Professors: Do Winners Get the Recognition that They Have Been Given?" *Psychology of Women Quarterly* 5 (5_suppl):767-772.
24. Key, Ellen, and Phillip Ardoin. 2019. "Students rate male instructors more highly than female instructors. We tried to counter that hidden bias." *Washington Post*, 8/20/2019. Accessed 9/3/2019. <https://www.washingtonpost.com/politics/2019/08/20/students-rate-male-instructors-more-highly-than-female-instructors-we-tried-counter-that-hidden-bias/>.
25. Kierstead, Diane, Patti D'agostino, and Heidi Dill. 1988. "Sex role stereotyping of college professors: Bias in students' ratings of instructors." *Journal of Educational Psychology* 80 (3):342.
26. Leslie, Sarah-Jane, Andrei Cimpian, Meredith Meyer, and Edward Freeland. 2015. "Expectations of brilliance underlie gender distributions across academic disciplines." *Science* 347 (6219):262-265.
27. MacNeill, Lillian, Adam Driscoll, and Andrea N Hunt. 2015. "What's in a name: Exposing gender bias in student ratings of teaching." *Innovative Higher Education* 40 (4):291-303.
28. Mengel, Friederike, Jan Sauermann, and Ulf Zölitz. 2018. "Gender bias in teaching evaluations." *Journal of the European Economic Association* 17 (2):535-566.
29. Miles, Patti, and Deanna House. 2015. "The Tail Wagging the Dog: An Overdue Examination of Student Teaching Evaluations." *International Journal of Higher Education* 4 (2):116-126.
30. Miller, JoAnn, and Marilyn Chamberlin. 2000. "Women are teachers, men are professors: A study of student perceptions." *Teaching Sociology* 28 (4):283.
31. Murray, Harry G. 1984. "The impact of formative and summative evaluation of teaching in North American universities." *Assessment and evaluation in Higher Education* 9 (2):117-132.
32. Peterson, David AM, Lori A Biederman, David Andersen, Tessa M Ditonto, and Kevin Roe. 2019. "Mitigating gender bias in student evaluations of teaching." *PLoS one* 14 (5):e0216241.
33. Piatak, Jaclyn, and Zachary Mohr. 2019. "More gender bias in academia? Examining the influence of gender and formalization on student worker rule following." *Journal of Behavioral Public Administration* 2 (2).
34. Rivera, Lauren A, and András Tilcsik. 2019. "Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation." *American Sociological Review* 84 (2):248-274.
35. Rowden, Gail V, and Robert E Carlson. 1996. "Gender issues and students' perceptions of instructors' immediacy and evaluation of teaching and course." *Psychological Reports* 78 (3):835-839.
36. Sidanius, Jim, and Marie Crane. 1989. "Job evaluation and gender: The case of university faculty." *Journal of Applied Social Psychology* 19 (2):174-197.
37. Sinclair, Lisa, and Ziva Kunda. 2000. "Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me." *Personality and social psychology bulletin* 26 (11):1329-1342.
38. Smith, Bettje P, and Billy Hawkins. 2011. "Examining student evaluations of Black college faculty: Does race matter?" *Journal of Negro Education* 80 (2).
39. Sporeen, Pieter, Bert Brockx, and Dimitri Mortelmans. 2013. "On the validity of student evaluation of teaching: The state of the art." *Review of Educational Research* 83 (4):598-642.
40. Sprague, Joey, and Kelley Massoni. 2005. "Student evaluations and gendered expectations: What we can't count can hurt us." *Sex Roles* 53 (11-12):779-793.
41. Stark, P, and R Freishtat. 2014. "An evaluation of course evaluations. ScienceOpen." Center for Teaching and Learning, University of California, Berkeley. Retrieved from <https://www.scienceopen.com/document>.
42. Storage, Daniel, Zachary Horne, Andrei Cimpian, and Sarah-Jane Leslie. 2016. "The frequency of "brilliant" and "genius" in teaching evaluations predicts the representation of women and African Americans across fields." *PLoS one* 11 (3):e0150194.
43. Subtirelu, Nicholas Close. 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.
44. Utti, Bob, Carmela A White, and Daniela Wong Gonzalez. 2017. "Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related." *Studies in Educational Evaluation* 54:22-42.
45. Utti, Bob, Carmela A White, and Alain Morin. 2013. "The numbers tell it all: students don't like numbers!" *PLoS one* 8 (12):e83443.
46. Wachtel, Howard K. 1998. "Student evaluation of college teaching effectiveness: A brief review." *Assessment & Evaluation in Higher Education* 23 (2):191-212.
47. Wallisch, Pascal, and Julie Cachia. 2019. "Determinants of perceived teaching quality: The role of divergent interpretations of expectations."
48. Wigginton, Henry, Nona Tollefson, and Edme Rodriguez. 1989. "Students' ratings of instructors revisited: Interactions among class and instructor variables." *Research in Higher Education* 30 (3):331-344.