

PANEL: Gender bias in student evaluations of teaching

Dr. Agnes G d'Entremont, University of British Columbia, Vancouver

Dr. Agnes d'Entremont is an Instructor in the Department of Mechanical Engineering. Her technical research in Orthopaedic Biomechanics is focused on joint motion and cartilage health with a particular concentration in pediatric hip disorders and MRI-based methods. Her teaching-related interests include team-based learning and the flipped classroom, as well as diversity and climate issues in engineering education.

Dr. Hannah Gustafson, University of British Columbia

Hannah Gustafson earned her PhD in Mechanical Engineering from the University of British Columbia. Her research focus is biomechanics.

PANEL: Gender Bias in Student Evaluations of Teaching

Student evaluations of teaching (SET) are important in the university setting for determining tenure and promotion. In 2000, 88% of deans surveyed reported their colleges “always used” systematic student ratings to evaluate faculty teaching performance, and that number increased in 2010 to 94%¹. Given the widespread usage of SET, it is important to consider biases, such as those due to gender, that may influence these evaluations. In particular, engineering has a low proportion of women faculty; only 15.7% of US engineering faculty were women in 2015². Efforts to increase the number of women faculty may be influenced by any gender biases in SET, as career trajectories may be influenced by SET results. In addition, engineering has 78.6% male student body². The interaction between student gender and instructor gender may also influence SET. This paper will provide a brief overview of research into the effect of professor (and student) gender on the SET scores, and briefly explore recommendations for mitigating the effects of bias.

SET is intended to quantify teaching effectiveness. However, it is not clear that this is what SET solely (or even primarily) measures. For example, one study showed a weak but positive correlation between the students’ learning, as assessed with a standard test before and after the course, and the course evaluation scores. They noted that “the section with the lowest course evaluation score witnessed similar amounts of estimated learning gains as the sections with the highest course evaluations”³. While our paper is not intended as a review of research on the relationship between SET and teaching effectiveness, it should be noted that much of the work described here is premised on the idea that unintended factors (including gender) likely influence SET scores to varying degrees.

The effect of instructor gender on SET has primarily been explored by aggregating and analyzing the SET scores for female and male instructors. Some studies measured no differences or small differences in the scores for female and male professors⁴⁻⁶ while other studies found that students rated female professors lower than male professors^{7,8}. However, these investigations do not control for differences in teaching styles and teaching effectiveness, making it difficult to determine the influence of gender bias; that is, they are measuring whether or not differences exist between instructor gender groups, rather than whether or not those differences are due to gender bias⁹. A gender bias could be masked at an institution if women faculty had a higher average teaching effectiveness⁹.

A few studies have controlled for differences in teaching effectiveness when exploring possible gender bias in SET. One study did this by disguising two instructors’ gender identity when teaching an online course¹⁰. For an online introductory anthropology/sociology course, students were divided into six discussion groups. Two assistant instructors (one female and one male) each taught two sections, one as their own identity and one as the other instructor’s identity. Therefore, four groups were included in the study, two groups in which the instructor’s actual gender was correctly identified to the students and two groups in which the instructor’s gender was the opposite of what the students were told. Efforts were made to ensure consistency in grading and the timing of feedback (neither faster nor slower to respond) between sections. When the SET results were grouped by actual instructor gender (i.e. the female instructor, whether the students believed she was female or male), no difference in the student ratings of the

instructors was found. However, when SET results were grouped by perceived instructor gender (i.e. both the female and male instructors, when the students believed each was female), students rated the perceived male instructor as significantly better than the perceived female instructor. These findings support the idea that there is a real bias that exists among students in evaluating instructors, not simply a difference in the teaching styles or teaching effectiveness between female and male instructors.

Another study performed a laboratory experiment where students were shown an identical lecture delivered by a stick figure with a gender-neutral voice and later were provided a written description of this professor (either female or male, either under 35 years old or over 55 years old) when filling out a teaching evaluation form¹¹. Student rankings for the professor, when identified as male, indicated they thought “he” was more enthusiastic, more frequently used voice tone to identify important concepts, was more likely to make students feel accepted and included, and showed more interest in the subject than the same professor when identified as female. Interestingly, there was no difference between male- and female-identified professors on content items on the same evaluation, namely “the ability to organize and logically deliver, with appropriate scientific terminology, precise lectures”¹¹.

Student ratings of male and female instructors are likely influenced by the gendered expectations of the rater¹². Two straightforward examples of gendered expectations are that students are more likely to assume a male instructor holds a PhD as compared with a female instructor¹³ and are more likely to refer to male instructors as “Doctor” or by their last name when writing evaluations of their instructors¹⁴. Each of these examples shows that the status and credentials of female faculty are discounted. Female professors may be expected to be more caring and nurturing than male professors, given that these traits are considered more feminine¹². Even though female instructors had more meetings with students outside of class hours, students were more likely to say female instructors were insufficiently available compared to male instructors⁵. Finally, enthusiasm (which can alone raise SET scores) is more consistent with a male gender stereotype, which may cause students to dismiss evidence of enthusiasm in female instructors (a form of confirmation bias), leading to lower SET scores for women¹¹. These all indicate different expectations for female and male instructors, in both their background and their educational style.

Personal experience may reduce students’ gendered expectations for female instructors, and thus reduce the influence any preexisting stereotypes have on SET. One study demonstrated that exposure to previous women instructors has been shown to reduce the bias in SET scores rating a single video lecture¹⁵. Other authors have theorized that students who do not interact much with an instructor rely more heavily on stereotypes for judgement rather than relying on the instructor’s actual behaviour¹³. In engineering, one challenge in applying these ideas is that there are many fewer opportunities for students to have personal experience learning from a female instructor, due to the low percentage of women faculty. On the other hand, women academics are perceived as exceptions, a perception that is heightened when they are in heavily male-dominated fields. This can translate into being perceived as exceptions to gender stereotypes as well, reducing the dissonance of women faculty acting in traditionally masculine ways⁵.

The gender of the student doing the rating, and the interaction between student and professor gender, may also influence SET ratings. Female students are more likely to choose a female instructor as their best instructor while male students were more likely to choose a male instructor as their best instructor¹⁶. Within the same class, female students gave female instructors higher SET scores than male students, but there were not significant differences between how male and female students rated their male instructors⁶. In another study, male students rated female instructors substantially lower than male instructors on all teaching evaluation measures, while female students rated female instructors somewhat lower on two-thirds of the measures⁷. If there are few female students, as is the case in most engineering classes², then any benefit of higher (or even not quite as reduced) female student ratings accrued to female instructors may be overbalanced by a preponderance of lower ratings from male students.

It is also important to acknowledge that other factors unrelated to either teaching effectiveness or gender also influence SET scores, including the students' grade expectations¹⁷, instructor race¹⁸, instructor age¹¹, and instructor charisma¹⁹.

In conclusion, there is evidence that gender bias influences SET, and that scores may be impacted by both instructor gender and student gender, with women faculty typically receiving lower scores. Caution is called for among engineering educators and administrators when using SET for advancement, tenure, and other decisions, due to the potential for compounding effects of even small gender biases on SET scores to limit the advancement of women faculty. Some recommendations for addressing the challenges of gender bias in SET include:

- Evaluating professors for their teaching in multiple ways to control for the acknowledged biases regarding gender. Beyond student evaluations, other tools for evaluating instructors include teaching portfolios, peer evaluations, outcome-based measures or discussions with groups of students regarding the professor's performance^{20,21}.
- Separating or removing survey items on instructor expressiveness, such as enthusiasm, warmth, confidence, and voice tone; SET may not accurately measure these, as they are filtered through existing biases about gender norms¹¹.
- Raising awareness of existing gender biases through workshops for professors or by instructors engaging with students about gender issues in the classroom²¹.
- Continuing research on gender biases in SET, particularly in engineering, where the majority of professors and students are male.

Questions for the panel:

- From your experiences, do you see evidence of different student expectations for female and male engineering professors, particularly related to attributes included in SET?
- Is bias (due to gender or other factors) in SET something that is acknowledged or addressed in your department? If so, how?
- How is SET used at your institution (e.g. for tenure/promotion?), and what is the potential impact of SET bias on the advancement of women engineering faculty?
- Engineering student bodies are typically predominantly male, as are engineering faculty rosters – what influence do you think this has on SET in engineering generally, and gender bias in SET specifically?

- What impact (positive or negative) do male colleagues' attitudes and actions have on the content of SET surveys, gender bias in SET, and the larger question of the advancement of women engineering faculty?
- What can institutions and individuals do to address and mitigate the effect of gender bias in SET, particularly within engineering?

1. Miller EJ, Seldin P. Changing Practices in Faculty Evaluation. *Academe*. 2014;100(3):35.
2. Yoder BL. Engineering by the Numbers. *ASEE*. 2015.
3. Beleche T, Fairris D, Marks M. Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. *Econ Educ Rev*. Elsevier Ltd; 2012;31(5):709–19.
4. Elmore PB, LaPointe KA. Effects of teacher sex and student sex on the evaluation of college instructors. *J Educ Psychol*. 1974;66(3):386–9.
5. Bennett SK. Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *J Educ Psychol*. 1982;74(2):170.
6. Centra JA, Gaubatz NB. Is There Gender Bias in Student Evaluations of Teaching? *J Higher Educ*. 2000;71(1):17–33.
7. Basow SA, Silberg NT. Student evaluations of college professors: Are female and male professors rated differently? *J Educ Psychol*. 1987;79(3):308–14.
8. Sidanius J, Crane M. Job evaluation and gender: The case of university faculty. *J Appl Soc Psychol*. 1989;19(2):174–97.
9. Feldman KA. College Students' Views of Male and Female College Teachers: Part I: Evidence from the Social Laboratory and Experiments. *Res High Educ*. 1992;33(3):317–75.
10. MacNeill L, Driscoll A, Hunt AN. What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. *Innov High Educ*. 2015;40:291–303.
11. Arbuckle J, Williams BD. Students' Perceptions of Expressiveness: Age and Gender Effects on Teacher Evaluations. *Sex Roles*. 2003;49(9–10):507–16.
12. Sprague J, Massoni K. Student Evaluations and Gendered Expectations: What We Can't Count Can Hurt Us. *Sex Roles*. 2005;53(11–12):779–93.
13. Miller J, Chamberlin M. Women Are Teachers, Men Are Professors: A Study of Student Perceptions. *Teach Sociol*. 2000;28(4):283–98.
14. Terkik A, Prud'hommeaux E, Alm CO, Homan C, Franklin S. Analyzing Gender Bias in Student Evaluations. Rochester Institute of Technology; 2016.
15. Fandt PM, Stevens GE. Evaluation Bias in the Business Classroom: Evidence Relating to the Effects of Previous Experiences. *J Psychol*. 1991;125(4):469.
16. Basow SA. Best and Worst Professors : Gender Patterns in Students' Choices. *Sex Roles*. 2000;43(5/6):407–17.
17. Boring A, Ottoboni K, Stark PB. Student evaluations of teaching (mostly) do not measure teaching effectiveness. *Sci Res*. 2016;1–11.
18. Bavishi A, Madera JM, Hebl MR. The effect of professor ethnicity and gender on student evaluations: Judged before met. *J Divers High Educ*. 2010;3(4):245–56.
19. Shevlin M, Banyard P, Davies M, Griffiths M. The Validity of Student Evaluation of Teaching in Higher Education: Love me, love my lectures? *Assess Eval High Educ*.

- 2000;25(4):397–405.
20. Stark PB, Freishtat R. An evaluation of course evaluations. Cent Teach Learn Univ California, Berkley. 2014;
 21. Laube H, Massoni K, Sprague J, Ferber AL. The Impact of Gender on the Evaluation of Teaching: What We Know and What We Can Do. NWSA J. 2007;19(3):87–104.