The Gender Bias in Student Evaluations of Teaching (SET) Literature in a Nutshell

1. Mengel, Friederike, Jan Sauermann and Ulf Zölitz, "Gender Bias in Teaching Evaluations, **Journal of the European Economic Association**, https://doi.org/10.1093/jeea/jvx057, April 2019.

   This paper provides new evidence on gender bias in teaching evaluations. The authors exploit a quasi-experimental dataset of 19,952 student evaluations of university faculty in a context where students are randomly allocated to female or male instructors. Despite the fact that neither students' grades nor self-study hours are affected by the instructor's gender, the authors find that women receive systematically lower teaching evaluations than their male colleagues. This bias is driven by male students' evaluations, is larger for mathematical courses, and is particularly pronounced for junior women.

2. Boring, Anne, "Gender biases in student evaluation of teaching," **Journal of Public Economics**, https://doi.org/10.1016/j.jpubeco.2016.11.006, January 2017.

   This article uses data from a French university to analyze gender biases in SETs. The university requires first year undergraduate students to take six mandatory courses, so students do not select their courses when they register. Students' assignment to male and female professors is random. The administration makes students' online ratings of professors mandatory. As all students across all sections of a discipline take the same final exam, it is possible to compare student learning at the end of the semester. The author studies whether a match between student and professor gender has an impact on a professor's overall rating. Gender biases appear to exist: male students give significantly higher ratings to male professors than to female professors. Male students also rate male professors significantly higher than how female students rate both female and male professors. Male students are more likely to give *excellent* ratings to male professors. The author also finds that students perform equally well on final exams whether their professor was a man or a woman, suggesting no difference in actual teaching effectiveness. Thus, the results suggest that differences in teaching skills are not driving gender differences in evaluations. As universities use SETs to decide on promotions and contract renewals, these results imply that promotion and hiring in universities may be biased against women. These biases may also be harmful to female students, given the main results from the literature that discusses the impact of a role model effect on student performance.

3. MacNell, Lillian, Adam Driscoll and Andrea Hunt, "What's in a name: Exposing gender bias in student ratings of teaching," **Innovative Higher Education**, https://doi.org/10.1007/s1075, August 2015.

   While difficult to separate gender from teaching practices in person, it is possible to disguise an instructor's gender identity online. In the experiment this article describes, 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. The difference between how students rated the two perceived genders stands in stark contrast to the fact that neither the actual male nor actual female instructor received significantly higher ratings than the other. Both instructors performed equally well from the students' perspective. However, in both cases the same instructor received different ratings depending solely on their perceived gender. In other words, when the actual male instructor was perceived to be female, he received significantly lower ratings than when he was perceived to be a male. For example, when the actual male and female instructors posted grades after two days as a male, this was considered by students to be a 4.35 out of 5 level of promptness, but when the same two instructors posted grades at the same time as a female, it was considered to be a 3.55 out of 5 level of promptness. In each case, the same instructor, grading under two different identities, was given lower ratings half the time with the only difference being the perceived gender of the instructor. Similarly, students rated the perceived female instructors an average of 0.75 points lower on the question regarding fairness, despite both instructors utilizing the same grading rubrics and there being no significant differences in the average grades of any of the groups. Despite the fact that the students were equally satisfied with the promptness and fairness of the actual instructors, the instructor that students perceived to be male was considered to be more effective. All this suggests that a female instructor would have to work harder than a male to receive comparable ratings. If female professors and instructors are continually receiving lower evaluations from their students for no other reason than that they are women, then this particular form of inequality needs to be taken into consideration as women apply for academic jobs and come up for promotion and review.

4. Boring, Anne, Kellie Ottoboni and Philip Stark, "Student evaluation of teaching (mostly) do not measure teaching effectiveness," **Science Open**, https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1, January 2016.

   The authors show the following: (1) SET are biased against female instructors by an amount that is large and statistically significant. (2) The bias varies by discipline and by student gender, among other things. (3) It is not possible to adjust for the bias, because it depends on so many factors. (4) SET are more sensitive to students' gender bias and grade expectations than they are to teaching effectiveness. (5) Gender biases can be large enough to cause more effective instructors to get lower SET than less effective instructors. The authors conclude: Overall, SETs disadvantage female instructors. There is no evidence that this is the exception rather than the rule. Hence, the onus should be on universities that rely on SET 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.

5. Arbuckle, Julianne and Benne Williams. “Students’ Perceptions of Expressiveness: Age and Gender Effects on Teacher Evaluations, **Sex Roles**, https://doi.org/10.1023/A:1025832707002 November 2003.

   The authors investigate the relationship between college students' perceptions of professors' expressiveness and implicit age and gender stereotypes. Three hundred and fifty-two male and female students watched slides of an age- and gender-neutral stick figure and listened to a neutral voice presenting a lecture, and then evaluated it on teacher evaluation forms that indicated 1 of 4 different age and gender conditions (male, female, “old,” and “young”). Main and interaction effects indicated that students rated the “young” male professor higher than they did the “young” female, “old” male, and “old” female professors on speaking enthusiastically and using a meaningful voice tone during the class lecture regardless of the identical manner in which the material was presented.