

Are Teaching Evaluations Sexist?



by Buce Gilley

In August, the American Association of University Professors (AAUP) filed an [amicus brief](#) alongside the faculty union of Nevada public universities in support of the gender-discrimination claims of Alice Wieland, a former University of Nevada, Reno, business professor. Wieland, whose research is on gender discrimination, claimed that her tenure committee based its decision on her poor teaching evaluations from students, which reflected not her teaching but widespread sexism against women. The lower court tossed the case out on summary judgement. Wieland appealed, and the AAUP jumped in.

In the brief, the AAUP insists that “a sizeable corpus of empirical research demonstrates that gender bias tends to

affect student evaluations of teaching.” This “well-established body of research,” it continued, “demonstrates that women and other marginalized groups face significant biases in student evaluations of teaching in higher education.”

As any academic knows, student evaluations are only one part of the teaching record that faculty present when they are up for promotion. Many other aspects are included, such as the rigor and innovativeness of their courses, a willingness to step into teaching gaps or develop new courses, and the taking on of grunt introductory courses with large enrollments. Student evaluations come in as a final check to make sure that the faculty member is delivering content and meeting his or her side of the bargain in areas like showing up for class, being prepared, grading in a timely manner, meeting with students, being well organized, and, yes, being pleasant and enjoyable to learn from. In my experience, those are the comments that committees pay attention to in reviewing student evaluations.

Before delving into the “well-established” and “sizeable” evidence that AAUP promises in its brief, pause for a moment to consider the implications of its argument that student evaluations are biased. Since there is an obvious observational problem (is a given female professor a poor teacher or a victim of discrimination?), any faculty member who could claim victim status could be automatically accorded an opt-out of student evaluations. Every time a female faculty (or member of another “marginalized group”) bombed her student evaluations, she could cite the “well-established” evidence. Indeed, a man who got poor teaching evaluations could, in the *zeitgeist* of today’s Left, come out as non-binary and claim transphobic bias.

These arguments would be especially useful for black faculty if they were not already near-guaranteed promotion due to the affirmative action-based obsessions of the contemporary

academy. An edited 2021 [collection](#) entitled *Implications of Race and Racism in Student Evaluations of Teaching* asserts widespread racial bias against black faculty and includes a chapter with the title (I am not making this up) “Dismantling the architecture of good teaching.”

Since the evaluation system would become useless if AAUP’s argument were to prevail, the only rational response of any department would be to exclude evaluations altogether, even for faculty in teaching-intensive roles. Teaching would be evaluated only using other criteria (quality of syllabi, third-party peer observations, etc.). Yet many women and others achieve promotion precisely by being popular teachers. The AAUP may be forcing more out than keeping more in with its advocacy.

What if, to take a second issue, there really are differences in how members of different groups comport themselves in the classroom? Here is the nub, because liberals never want to admit the possibility of such differences, even as they promote identity categories with such fervor. Could it be that men and women differ in their social conceptions of what they should do in the classroom, which might lead to systematically different levels of teaching effectiveness?

Cue the “mountains of evidence” to the contrary. I will return to that below. Even if there is no “objective” evidence of gender (or racial) differences in teaching quality, we are still left to explain the *subjective* differences without reaching for the easy explanation of discrimination. To take gender, for example, assume that half the students in the class are males who prefer to learn from males, whereas females are neutral. If so, female faculty will earn systematically lower evaluations. Sexism! But wait: Isn’t a core tenet of the “cultural competency” and “diverse faculty” ideologies that some students learn better when they see someone like themselves at the front of the classroom? If this is true of males, then why should they be denied reflections

of their own selves as others are afforded them?

So what exactly is the AAUP advocating? That students be forced to learn from people who they cannot relate to because, well, the problem is theirs? To play devil's advocate, students are paying the bills and should have the choice to learn from whomever they please. If they are forced to study under someone whom they find less conducive to their learning, why should they not express that? If the AAUP was concerned with education rather than social engineering, it would recognize the autonomy of universities and departments to hire professors that students will be eager to learn from, full stop.

Further, what would be the negative effects on teaching on a campus with no accountability mechanisms at all? In other words, what sorts of "bias" would be introduced into the university classroom if students had little or no say in evaluating their instructors? Why should *faculty* biases about whom to put in the classroom outweigh *student* biases about whom to learn from?

This in turn raises an intriguing possibility: If the AAUP is going to go to the floor and insist that male students who evaluate female professors worse than male professors should have their evaluations removed for gender bias, would that not also apply to female (or black) students who evaluate male and white professors worse? And, given the lack of observational evidence, would that mean that, say, a male professor in a female-dominated discipline (like psychiatry or art history) who was denied tenure because of teaching evaluations would have a *prima facie* case against his department? Cue another AAUP amicus brief?

All of this is mere prelude to the fact that the evidence cited in the AAUP brief is as crumbly as a wedding cake. So the arguments above may not actually be needed to defend student evaluations.

In a classic example of “policy-based evidence-making,” the AAUP digs up three pieces of research that allegedly prove the case, while ignoring a significant amount of research that shows otherwise. Let’s examine each one.

The [first](#) is a 2021 article entitled “Evaluating Student Evaluations of Teaching: A Review of Measurement and Equity Bias in SETs and Recommendations for Ethical Reform” by Rebecca Kreitzer and Jennie Sweet-Cushman. It is described by the AAUP as “a prominent 2021 metastudy of more than 100 articles.” But it is nothing of the sort. It is a literature review, not a metastudy (which pools data in a statistical manner). The main intention of the authors is to use their literature review to challenge a 2012 [metastudy](#) that found no gender bias. But they have nothing to say about the methods of that study, only that they don’t agree with it.

In any case, the authors make clear that the literature is mixed, including many studies that find “women to be advantaged in evaluations, especially in departments where women are overrepresented, such as certain humanities fields.” They also make clear an important aspect of gender bias: It is usually not against females per se but against females who do not align with the expectations of female behavior. In other words, since females in academia are far less representative than females in society at large (unlike males), they may be graded lower due to gender norms. We may or may not think that’s bad, but it certainly highlights another serious problem: the unrepresentative nature of female academics compared to females in society at large.

The AAUP alleges that a [second](#) source cited in its brief shows bias in economics course evaluations. But again, the actual findings are something else. The paper finds that, at the beginning of a course, there is no gender bias in student evaluations. Students rate male and female professors about equally. As the course progresses, however, the ratings for females stay the same, while the ratings for men go up: “We

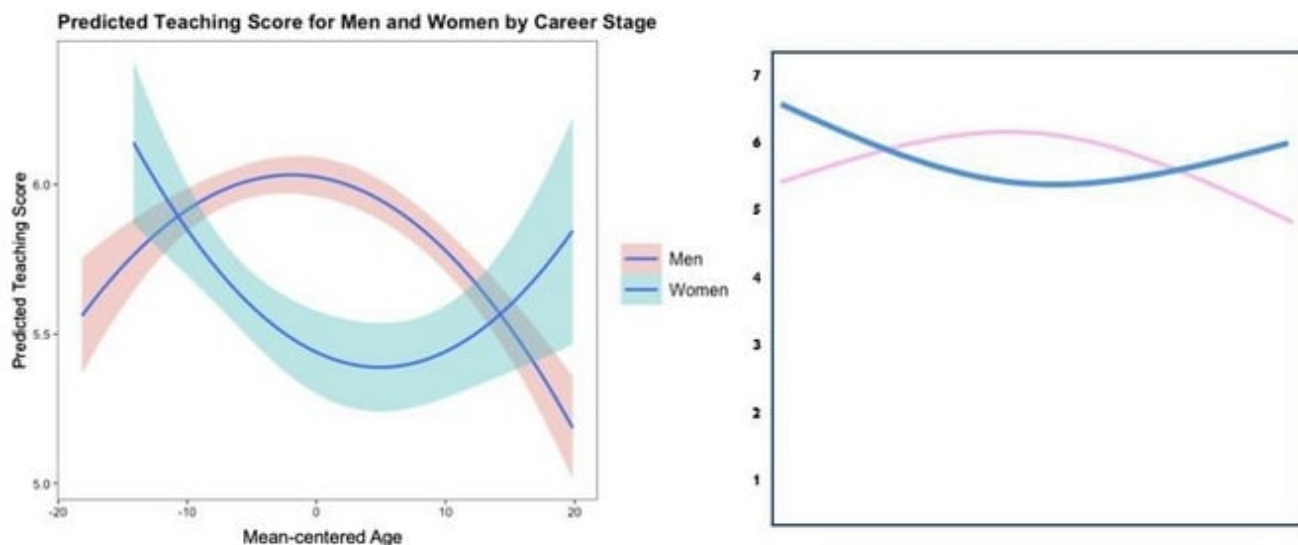
see that men increase in their ratings for all characteristics between Time 1 and Time 2 indicating that students see men more favorably as time goes on, which does not happen for women.” This is a significant wrinkle in the claim that this study finds bias “against women.”

Maybe students begin with an open mind and then generally find that they are learning more from male instructors. Or perhaps they assume, as they read more and more key research with male authors, that they are getting a better education from a male professor. Of course, we should urge students to be less prone to these cognitive shortcuts. It may not be until many years later that they realize which faculty “really had an impact on me.” But, after all, such young people are called “students” for a reason. Shall we bulldoze their instincts in pursuit of a threadbare case alleging sexism?

The [third](#) piece of research cited by AAUP is the most egregious junk of all. In the article “Agentic But Not Warm: Age-Gender Interactions and the Consequences of Stereotype Incongruity Perceptions For Middle-Aged Professional Women,” five scholars examined 59,600 student evaluations of faculty teaching in an MBA program. On a 1 to 7 scale, the average overall evaluation score for male teachers was 5.87, whereas for female teachers it was 5.62, a 4-percent difference. Given that it takes a serious degree of poor evaluations for such scores to matter to a committee (in my experience at least two points on a seven-point scale, though more likely three or four), this difference is likely *inconsequential*.

The real finding of the study is that males and females differ systematically in how their evaluations change from early- to mid- to late-career stages. Men start lower, excel in mid-career, and then drop off. Women start higher, tank in mid-career, and then regain their mojo. The study “reveals a significant decline in women’s teaching evaluations from young adulthood to middle age (and a rebound from middle age to older adulthood)” while men “increase from young adulthood to

middle age” and then decline. I reprint their estimation of these trends below on the left.



Does this show gender bias against women? No, even if we accept the evaluations as having objective value. Instead, they merely show that students are sexist against early-career men, mid-career women, and late-career men.

In any case, this alleged sexism, like the average differences, is very small. The graphic they produce exaggerates the size of the differences. The Tufte Lie Factor test (a measure of graphical misrepresentation) gives the authors’ graphic a score of seven, whereas a “fair” graphic should never exceed two or three. I produce a graphic to the right of the original with a Tufte Lie Factor of one (no distortion) to show the difference.

Moreover, this study puts an elephant in the room and then tiptoes around it. What if those small shifts in evaluations reflect some underlying *reality* about teaching performance across career spans? The authors blithely assert that they “know of no theory” that could explain why women might become less effective or likeable teachers in middle age. No theory? Marriage, children, physical and mental changes, dashed hopes? The literature is actually vast.

The AAUP's crusade against teaching evaluations echoes another contemporary crusade in the academy against research-citation counts. These were originally embraced by feminists because they would reveal the stellar research impact of hitherto obscure female academics and how they had been systematically discriminated against by their male colleagues. Instead, such counts often showed that males were on average [more productive](#), both in research outputs and impacts.

The response of the academy? Citations are themselves biased and should be either dismissed or [remade](#) so that faculty are forced to cite black women or use a gender-balance citation tool, to name two related projects.

By the way, Alice Wieland's citation count in Google Scholar, after more than a decade of research, was 638 at last count, which would not be impressive to any promotion committee. A recently promoted female business faculty member at University of Nevada, Reno (the same department where Wieland worked), Jinyu Hu, has 2,803 citations, while a recently promoted male in the department, Charles Carslaw, has 1,531.

Given Wieland's thin research record, the committee would have put a lot of store on her teaching. If Wieland's teaching evaluations were deemed insufficient, the best guess is that she was an insufficient teacher and was appropriately denied promotion.

Bruce Gilley is a professor of political science at Portland State University and the author, most recently, of The Case for Colonialism.