Is Gender Bias an Intended Feature of Teaching Evaluations?

Such evaluations pretend to be the result of a neutral process but are better measures of student stereotypes than teaching effectiveness, argues Victor Ray.

By Victor Ray, Inside Higher Ed, February 9, 2018

Every semester brings with it a new series of articles, blog posts and stories about gender and racial biases in teaching evaluations. A large and constantly growing body of academic literature demonstrates how bias shapes these tools.

For example, experiments with students in online courses show that identical courses are rated lower if the instructor is randomly assigned a woman’s name. Students may also use evaluations to comment on faculty appearance, tone of voice or even their sexual orientation.

And in a political environment where much of the population denies basic empirical facts about racial and gender inequality, teaching about such controversial subjects can open one up to claims of political bias. Social sciences don’t have laws, but if we were attempting to devise one, “Women and minorities get lower teaching evaluations” would be pretty close to axiomatic.

From a methodological standpoint, teaching evaluations are a mess. These evaluations lack external validity and don’t correlate with student learning outcomes. Typically, when social scientists recognize a research instrument is providing an incorrect measure, or that a measure is systematically biased, the measure is abandoned and (hopefully) replaced with a better one. Everyone knows -- or should know -- that teaching evaluations are better measures of student stereotypes than teaching effectiveness. Yet colleges and universities persist in laundering systematic bias through tenure and promotion processes, the legitimacy of which depend upon their supposed neutrality.

Although the methodological problems with these tests matter, it is also important to not get lost in the abstract; we must remember that biased evaluations can actually destroy people’s dreams. Promotions, raises and tenure are partially based on biased evaluations. Students who are unhappy with a grade, who dislike the opinions of a disciplinary expert or who are simply sexist can play an outsize role in their instructor’s future job negotiations.

Using biased evaluations allows colleges and universities to punish those whose identities deviate from white male normativity. Take a hypothetical gender discrimination case for denial of tenure because of poor teaching. Substantiating the harm requires evidence that discrimination was based on membership in a protected category. The institution will be able to point to poor teaching evaluations -- despite their known biases -- to argue that denial of tenure was based on less meritorious teaching.

The irony here is clear; the discriminatory bias built into measures of teaching effectiveness can be retroactively used to justify unequal outcomes based on those measures. Biased evaluation criteria explain biased outcomes, which the college or university then considers legitimated. The case of Ellen Pao followed just this logic after she filed a discrimination suit against a venture capital firm in Silicon Valley. A jury saw her performance evaluations as legitimate, despite the fact that they were systematically lower once she complained of gender discrimination -- a pattern she attributed to retaliation.

Feminist sociologists have long argued that one of the features of contemporary organizations is their gendered nature. Claiming that organizations are gendered means that supposedly gender-neutral jobs are actually considered men’s prerogative and that women in “men’s jobs” -- like, say, professors -- are
thought of as interlopers, out of place or what consistently shows up on evaluations: less competent. The key here is that, of course, these ratings aren’t based on objective measures of competence; rather, they are sifted through widely held stereotypes about women. But they are given the patina of legitimacy once the institution accepts them as credible when it comes to retention, promotion and tenure. That also has the benefit of protecting the university from potential lawsuits down the road.

Biased teaching evaluations, like race- and class-biased test scores, support the status quo and don’t create the same type of public outcry as, for instance, certain affirmative action policies that may slightly disadvantage white men. They are the perfect vehicle for a type of gender-blind discrimination because they allow one to claim detachment and objectivity. They pretend the “best qualified” is measured and confirmed through a neutral process that just so happens to confirm the worst stereotypes about women. Recent research by Katherine Weisshaar shows that, even accounting for productivity, gendered differences in the way tenure committees evaluate women contributes to fewer women moving up.

Given the evidence, I’ve almost reached the conclusion that gender- and race-biased evaluations are not used in spite of their bias, but because they are biased. Of course, my assumption about bias being a desirable feature of student evaluations is speculative. Hard evidence would be difficult to establish, because part of the point of these evaluations is that the bias is plausibly deniable. Administrators are unlikely to admit that they knowingly employ a biased instrument in employment proceedings. But I think the evidence for concluding that bias is a feature of teaching evaluations is in my favor. If colleges and universities know that evaluations are biased, and those biases are more likely to harm women and minorities (and the intersection of these categories), at what point can we start to assume that such evaluations are an intended feature of the process?

I don’t believe colleges and universities are going to drop these types of evaluations anytime soon, as the pressure to quantify every area of academic life is increasing. But making department heads, deans and tenure committees aware of both the biased nature of these evaluations and how they can influence tenure decisions will help to reduce the harm that reliance on such biased measure inflicts.

Victor Ray is an assistant professor of sociology at the University of Tennessee at Knoxville. His research examines race and gender discrimination in organizations. His commentary has appeared in Newsweek, Boston Review and Gawker. He is the new editor of “Conditionally Accepted,” an Inside Higher Ed career advice column for marginalized academics.