'Lovely' and 'scientific'—Medical student evaluations differ by gender and minority status

April 16 2019

In the largest analysis to date of narrative medical school evaluations, researchers at UC San Francisco and Brown University have found significant differences in how female and underrepresented minority medical students are described.

The researchers used natural language processing to analyze a large bicoastal sample of nearly 90,000 narrative evaluations of third-year clerkships from UCSF and Brown. The data spanned nine years at UCSF—from 2006 to 2015—and five at Brown—from 2011 to 2016.

These evaluations are supposed to focus on student behaviors—or competencies—that are directly relevant to medicine. But the analysis found that evaluators often used personal descriptors to describe a student's performance, and they used strikingly different words for men and for women.

The study also identified personal descriptors that were applied differently depending on whether students were members of groups that are underrepresented in medicine (URM). But nearly all of these personal descriptors were used more often to describe students who were not in those groups.

The evaluations form the basis of the grades that students get in their core clerkships, which are like medical apprenticeships, and are

frequently quoted in letters of recommendation for residencies. Even small biases can snowball, with lasting effects on students' career prospects.

"There shouldn't be systematic differences based on gender or URM status in a sample this big," said Urmimala Sarkar, MD, MPH, an associate professor of medicine at UCSF and the senior author of the study, published Tuesday, April 16, in the *Journal of General Internal Medicine*. "Everything should come out in the wash."

A fourth-year UCSF medical student, Alexandra Rojek, used natural language processing to find the words that were commonly used to describe students—yet not so common that they were used to describe too many different things and therefore could have a wide range of meanings—and that also were used more or less often depending on a student's gender or URM status.

The analysis found 37 descriptive words that evaluators applied differently by gender and 53 descriptive words applied differently by URM status.

"Looking at the words, we realized a lot of these terms were personality descriptors," Rojek said. "While one could say there are only 37 differences, this is after we filtered out words that don't appear very often. Even a handful of these words may be enough for us to be concerned."

The study authors determined that nearly two-thirds (23 of 37) of the words used differently by gender described students' personal attributes, and just over half of these (13 of 23) were more likely to be used to describe women.

These included "pleasant," which was associated with getting a passing

grade; "energetic," "cheerful," and "lovely," which were not associated with any particular grade; and "wonderful" and "fabulous," which were associated with honors grades.

The words used more commonly for men included "respectful" and "considerate," which had no association with a grade, while "good" was associated with earning passing grades, and "humble" was more common among those who earned honors grades.

Descriptive words that varied by gender also included four competency-related words. "Efficient," "comprehensive," and "compassionate," which the raters deemed related to competency in patient care, predominated for women, and were associated with honors grades. "Relevant" was more common for men, and was also associated with honors grades.

Of the 53 words that differed by URM status, nearly a third (16 of 53) described personal attributes, and the overwhelming majority (13 of 16) were used for non-URM students.

Evaluators were more likely to use words like "pleasant," "open," and "nice" to describe URM students, and these words were associated with passing grades. Descriptors like "enthusiastic," "sharp," and "bright" were used more commonly for non-URM students. These words were not associated with getting a particular grade, but "mature" and "sophisticated" were more frequently associated with honors grades.

Just over a quarter of the words (15 of 53) were related to competencies, and all of them were used more often for non-URM students. These included, "outstanding," "impressive," and "advanced," which were associated with honors, while "superior," "conscientious," and "integral" were not associated with a grade.

"Even when we think we're being thoughtful, we can fall into old tropes," said Catherine Lucey, MD, executive vice dean and vice dean for education at [UCSF School of Medicine](#). "We need to remind faculty there are words they should try to avoid—words they believe are positive, but are highly gendered or highly racially stereotyped. This paper opens up the door to how we can address this."

Provided by University of California, San Francisco