

THE CHRONICLE of Higher Education

TEACHING

Colleges Are Getting Smarter About Student Evaluations. Here's How.



Amanda L. Smith for The Chronicle

Bill Harbaugh, a professor of economics at the U. of Oregon: "It's pretty clear that if there's a number out there, it'll get misused."

By *Kristen Doerer* | JANUARY 13, 2019

✓ PREMIUM

Emily Wu and Kenneth Ancell, two students at the University of Oregon, approached their honors research professor, Bill Harbaugh, a few years ago about studying the relationship between student evaluations and grade inflation. Harbaugh, a professor of economics, was enthusiastic. Wu and Ancell dived into the university's extensive data on evaluation and transcripts, focusing on its two largest schools, journalism and business.

What they found surprised them.

"Having a female instructor is correlated with higher student achievement," Wu said, but female instructors received systematically lower course evaluations. In looking at prerequisite courses, the two researchers found a negative correlation between students' evaluations and learning. "If you took the prerequisite class from a professor with high student teaching evaluations," Harbaugh said, "you were likely, everything else equal, to do worse in the second class."

The team found numerous studies with similar findings. "It replicates what many, many other people found," said Harbaugh. "But to see it at my own university, I sort of felt like I had to do something about it."

He did. In the spring of 2017, Harbaugh assembled a task force on the issue and invited Sierra Dawson, now associate vice provost for academic affairs, to join. The last time that course evaluations had been reviewed was a decade earlier, when the university moved from a paper system to online.

Oregon is among a small but growing number of institutions that are looking to de-emphasize the use of student evaluations in personnel decisions. Instead, faculty members are increasingly evaluating one another's teaching. They are also writing reflections on their own teaching.

Meanwhile, even where student evaluations still play a significant role, they are being revised to minimize bias and prompt more-thoughtful feedback. The evaluations' wording puts greater emphasis on student engagement and the shared responsibility between instructor and student.

Studies since the 1980s have found gender bias in student evaluations and, since the early 2000s, have found racial bias as well. A 2016 study of data from the United States and France found that students' teaching evaluations "measure students' gender biases better than they measure the instructor's teaching effectiveness," and that more-effective instructors got lower ratings than others did.

In his 2003 book, *Grade Inflation* (Springer), Valen Johnson, a professor of statistics at Texas A&M University at College Station, argued that the onset of student evaluations had brought about rampant grade inflation, as professors realized they could "buy" better evaluations with easier grading. In a 2016 survey of faculty members by the American Association of University Professors, 67 percent said student evaluations put upward pressure on grades. Canadian researchers conducted a meta-analysis of 97 studies that "revealed no significant correlations between the ... ratings and learning."

Linda Nilson, director emerita of Clemson University's Office of Teaching Effectiveness and Innovation, said that several decades ago there was a moderate correlation between student ratings and student learning. Over the years it has disappeared.

Nilson and Peter F. Lake, director of the Center for Excellence in Higher Education Law and Policy at Stetson University, point to the changing relationship between students and their college or university. Students now are treated as customers, and their evaluations are a metric of satisfaction, not academic progress.

Despite the data, at many colleges, particularly research-based institutions, student evaluations are still the main measure, if not the only one, of teaching effectiveness in promotion-and-tenure decisions.

Seeking Alternatives

Some colleges, however, are taking the evidence to heart and reappraising the role that student evaluations play in their faculty members' careers. Along with Oregon, institutions seeking alternative methods include Colorado State University at Fort Collins, the University of Colorado at Boulder, the University of Kansas, the University of Massachusetts at Amherst, the University of Southern California, Ryerson University, in Toronto, and a division of the University of California at Berkeley.

Oregon's task force set out to help departments define excellence in teaching, establish resources to help instructors develop their skills, and offer detailed criteria for how instructors would be evaluated.

It identified three windows through which to evaluate teaching: students, peers, and professors themselves.

A year into the project, the task force discovered that Southern California was attempting a similar reform. The impetus to revise student evaluations there began with Michael Quick, the provost. When he read the scholarship on bias in student evaluations, he initially banned their use as a primary measure in promotion-and-tenure decisions, asking departments to use other metrics. The university has since clarified that stance, after faculty input, and has moved to a largely peer-review process, with student evaluations playing only a part in measuring student engagement.

An argument often made by faculty members is that students haven't been trained in pedagogy and can't give feedback on instruction based on best practices. Professors, by contrast, can. But academics have their own biases, so Southern California has based even the peer-review process on observable criteria and has required professors to receive anti-bias training. Peers are asked to evaluate instructors' teaching by observing classes, reviewing course materials, and considering instructors' written reflections.

"The criteria in those evaluation tools are observable, objective behaviors or characteristics of a course," said Ginger Clark, assistant vice provost and director of USC's Center for Excellence in Teaching. "So very little within the tools that we've created is subjective narrative."

If schools within the university want to use those evaluation tools, personnel from the teaching center will train their faculty members, she said. Alternatively, professors are welcome to use tools developed in their fields, but they must provide peer-review training to increase accuracy and decrease bias.

Oregon was already using peer evaluation, but to various degrees and with various levels of success, depending on the department. Now it is trying to elevate peer evaluation to a uniform high standard, and the faculty senate passed a measure to include instructors' reflections.

In the division of mathematical and physical sciences at Berkeley, department chairs and an ad hoc committee are instructed to read instructors' written reflections to see how they use evaluations to inform their teaching. Philip B. Stark, associate dean of the division, gives the example of students' complaining about an assigned textbook. If a professor says she is writing her own textbook because existing ones aren't very good, that provides helpful context for the committee to consider, he said.

Legal Pressure

Doing nothing to revise or phase out student evaluations could be a risky proposition not just educationally, but also legally.

In August, an arbitrator ruled that Ryerson could no longer use student evaluations to gauge teaching effectiveness in promotion-and-tenure decisions. The Ryerson Faculty Association brought the arbitration case and argued that because of the well-documented bias, student evaluations shouldn't be used for personnel decisions.

"This is really a turning point," said Stark, who testified on behalf of the Ryerson faculty group. He thinks the United States will see similar cases. "It's just a question of time before there are class-action lawsuits against universities or even whole state-university systems on behalf of women or other minorities, alleging disparate impact."

Ken Ryalls, president of the IDEA Center, a nonprofit higher-education consulting organization, recognizes the bias but thinks doing away with evaluations isn't the answer. He opposes efforts to eliminate the voice of students. "It seems ludicrous," he said, "to have the hubris to think that students sitting in the classroom have nothing to tell us."

"The argument that you should get rid of student evaluations because there is bias inherently is a bit silly," he said. "Because basically every human endeavor has bias."

The goal should instead be to minimize or eliminate bias, he argued. IDEA has been working on just that, and so far, Ryalls said, studies suggest that it is succeeding in finding ways to counter gender bias.

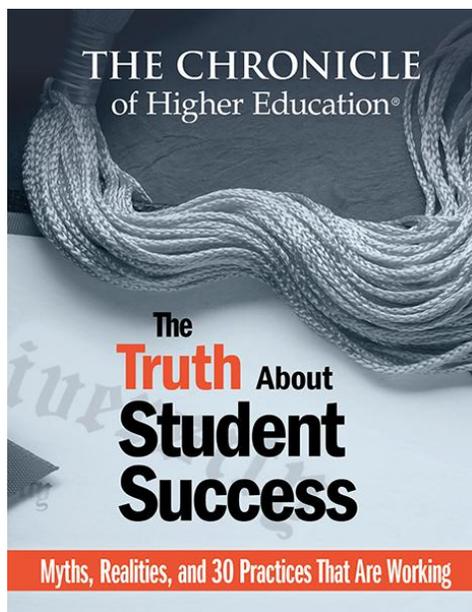
Most course evaluations have some generic questions, such as "Overall, how do you rate this instructor?" or "Overall, how do you rate this course?" The broadness of those questions opens up student evaluations to bias because they are "not tied to any particular instructor behavior," explained Clemson's Linda Nilson, who has observed IDEA's efforts.

IDEA offers sample questions about whether a student feels that learning outcomes have been achieved, about self-efficacy (did the student feel capable of succeeding?), about teaching methods the student observed, and about the student's motivation to take the course. Such questions not only are more specific but also say something about the challenges the professor faced, information that is weighted in IDEA's system.

Many questions also take some of the onus of student learning off the instructor and make it clear that it is a shared responsibility between students and instructors. Nilson thinks that principle could be emphasized even more.

Southern California has administered its own revised course evaluations twice and is about to look at the data again to see if more revisions are needed. Questions examine whether the course objectives were well explained, assignments reflected the material covered, and the instructor sufficiently explained difficult concepts, methods, and subject matter. The university hopes the specificity of the questions will minimize bias, but it has decided, in any case, that the evaluations will make up only a small portion of the teaching-evaluation portfolio.

The Truth About Student Success



The Truth About Student Success: Myths, Realities, and 30 Practices That Are Working

\$179.00

Print



ADD TO CART

The University of Oregon, which has students answer evaluation questions on a one-to-five scale, is looking to eliminate numerical ratings. "It's pretty clear that if there's a number out there, it'll get misused," said Harbaugh, the economics professor.

Oregon decided to have students select, from a list, teaching elements that were most beneficial to their learning and those that could use some improvement. They were then asked to provide written comments about those areas. The responses are aggregated, so professors can see if a cluster of comments indicates particular weaknesses or strengths.

The goal of all of those efforts is not only to minimize bias but also to ensure that instructors can learn from student feedback and act accordingly. "It's so important," said Stetson's Peter Lake, "not to weaponize student evaluations against people but to use them constructively."

What Works, What Doesn't

That's in large part why Oregon decided to try a midterm student-experience survey that only the applicable faculty member can view. An instructor can make changes in the middle of a semester, when students can still benefit, encouraging them to give constructive feedback.

"To be totally honest, I stopped looking at the numerical feedback 12 years ago, because it didn't mean anything," said Chuck Kalnbach, a senior instructor of management at Oregon, who started with the pilot program last spring. In contrast, he has found the midterm survey helpful, asking students about one thing that's working well and one that's not, with space to explain.

Kalnbach's organizational-development and change-management students said "transparency of instructions and grading" could use improvement. They wanted more direction and clarity on what was going to be on the midterm exam, which they had just taken. He had purposely not given them a study guide or offered much specific information on the midterm. After reading through the survey results, he explained to them that his "class is all about dealing with ambiguity," and that he wanted them to be able to deal with ambiguous and conflicting information. "Life," he said, "doesn't provide a study guide."

"That's information I know I can stress more," he said, "and when I teach the class this year, I'm going to stress that right upfront. I'm going to acknowledge that they're going to be frustrated, and I'm going to tell them it's part of the process."

That type of feedback proved to be popular among other professors, too, and the faculty senate voted to approve the midterm survey. Dawson, the associate vice provost at Oregon, and Harbaugh expect that the university will begin using it in the fall.

Oregon students like it too. Marlene H. Loui, a senior, appreciated how the new midterm survey and revamped end-of-term versions had made her think harder about why some teaching methods worked better than others. Usborn Ocampo, also a senior, was surprised to learn about the implicit bias in student evaluations. He said most students aren't familiar with the thinking behind the new evaluations, but he hopes that will change when the task force holds focus groups with them this winter.

Kate Myers, an instructor in Oregon's English department, found the numbers generated from the old survey so useless that she began distributing her own end-of-term survey. "Students would often just go down the line and hit all fives or all fours or whatever, without really thinking about it," she said. "I don't know what a student thinks a five-level class is or a four-level class. That doesn't make any difference to the way I teach my class, because I'm not getting substantive feedback."

The university's new questions focus on student engagement rather than the instructor, an approach she said is more helpful in considering her teaching methods.

The University of Washington, Oregon State University, and Portland State University have expressed interest in the University of Oregon's work.

"I think they're kind of waiting for us to see how it plays out," Harbaugh said. At a recent conference of the International Society for the Scholarship of Teaching and Learning, in Norway, the issue of student evaluations came up repeatedly, said Dawson. "Literally all over the world, people are trying to solve this problem."

In the meantime, even when evaluations are used, caveats are more often attached. In November, Oregon's faculty senate passed disclaimer language, noting that student evaluations are not generally reliable measures of teaching effectiveness, and that they are affected by documented gender, racial, and ethnic biases. That language will go into faculty members' promotion-and-tenure files in January.

After Berkeley's history department decided to switch from paper to online evaluations and held a discussion on the topic, Brian DeLay, an associate professor, tweeted that professors — especially white, tenured men — should clue in their students to the evaluations' bias.

DeLay doesn't think talking with students about the problem before they turn in evaluations will avoid the biases, which are societal. But he does think that students "deserve to know the truth about these evaluations," and that talking about it "helps us have this broader campuswide conversation."

A version of this article appeared in the January 18, 2019 issue.

© 2019 The Chronicle of Higher Education