

Teaching Eval Shake-Up

[IHE insidehighered.com/news/2018/05/22/most-institutions-say-they-value-teaching-how-they-assess-it-tells-different-story](https://www.insidehighered.com/news/2018/05/22/most-institutions-say-they-value-teaching-how-they-assess-it-tells-different-story)

Research is reviewed in a rigorous manner, by expert peers. Yet teaching is often reviewed only or mostly by pedagogical non-experts: students. There's also mounting evidence of bias in student evaluations of teaching, or SETs -- against female and minority instructors in particular. And teacher ratings aren't necessarily correlated with learning outcomes.

All that was enough for the University of Southern California to do away with SETs in tenure and promotion decisions this spring. Students will still evaluate their professors, with some adjustments -- including a new focus on students' own engagement in a course. But those ratings will not be used in high-stakes personnel decisions.

The changes took place earlier than the university expected. But study after recent study suggesting that SETs advantage faculty members of certain genders and backgrounds (namely white men) and disadvantage others was enough for Michael Quick, provost, to call it quits, effective immediately.

'I'm Done'

"He just said, 'I'm done. I can't continue to allow a substantial portion of the faculty to be subject to this kind of bias,'" said Ginger Clark, assistant vice provost for academic and faculty affairs and director of USC's Center for Excellence in Teaching. "We'd already been in the process of developing a peer-review model of evaluation, but we hadn't expected to pull the Band-Aid off this fast."

While Quick was praised on campus for his decision, the next, obvious question is how teaching will be assessed going forward. The long answer is through a renewed emphasis on teaching excellence in terms of training, evaluation and incentives.

"It's big move. Everybody's nervous," Clark said. "But what we've found is that people are actually hungry for this kind of help with their teaching."

SETs -- one piece of the puzzle -- will continue to provide "important feedback to help faculty adjust their teaching practices, but will not be used directly as a measure in their performance review," Clark said. The university's evaluation instrument also was recently revised, with input from the faculty, to eliminate bias-prone questions and include more prompts about the learning experience.

Umbrella questions such as, "How would you rate your professor?" and "How would you rate this course?" -- which Clark called "popularity contest" questions -- are now out. In are questions on course design, course impact and instructional, inclusive and assessment practices. Did the assignments make sense? Do students feel they learned something?

Students also are now asked about what they brought to a course. How many hours did they spend on coursework outside of class? How many times did they contact the professor? What study strategies did they use?

While such questions help professors gauge how their students learn, Clark said, they also signal to students that “your learning in this class depends as much as your input as your professor’s work.” There is also new guidance about keeping narrative comments -- which are frequently subjective and off-topic -- to course design and instructional practices.

Still, SETs remain important at USC. Faculty members are expected to explain how they used student feedback to improve instruction in their teaching reflection statements, which continue to be part of the tenure and promotion process, for example. But evaluation data will no longer be used in those personnel decisions.

Schools and colleges may also use evaluations to gather aggregate data on student engagement and perceptions about the curriculum, or USC’s diversity and inclusion initiatives, Clark said. They may also use them to identify faculty members who do “an outstanding job at engaging students, faculty who may need some support in that area of their teaching, or problematic behaviors in the classroom that require further inquiry.”

Again, however, SETs themselves will not be used as a direct measure in performance evaluations.

More Than a Number

While some institutions have acknowledged the biases inherent in SETs, many cling to them as a primary teaching evaluation tool because they’re easy -- almost irresistibly so. That is, it takes a few minutes to look at professors’ student ratings on, say, a 1-5 scale, and label them strong or weak teachers. It takes hours to visit their classrooms and read over their syllabi to get a more nuanced, and ultimately more accurate, picture.

Yet that more time-consuming, comprehensive approach is what professors and pedagogical experts have been asking for, across academe, for years. A [2015 survey](#) of 9,000 faculty members by the American Association of University Professors, for instance, found that 90 percent of respondents wanted their institutions to evaluate teaching with the same seriousness as research and scholarship.

The survey gave additional insight into the questionable validity of SETs: two-thirds of respondents said these evaluations create pressure to be easy graders, a quality students reward, and many reported low rates of feedback.

Echoing other studies and faculty accounts, responses to the AAUP survey suggested that SETs have an outsize impact on professors teaching off the tenure track, in that high student ratings can mean a renewed contract -- or not.

The AAUP committee leading the 2015 study argued that faculty members within departments and colleges -- not administrators -- should develop their own, holistic teaching evaluations. It also urged “chairs, deans, provosts and institutions to end the practice of allowing numerical rankings from student evaluations to serve as the only or the primary indicator of teaching quality, or to be interpreted as expressing the quality of the faculty member’s job performance.”

Faculty committees at USC also have worked to address teaching excellence for the past five years, recommending that the university invest more in teaching, adopt incentives for strong instruction, and move toward a peer model of review.

USC’s teaching evaluation plan reflects some of those recommendations -- as well as a new emphasis on teaching excellence.

“We must renew our focus on the importance of teaching and mentorship, putting into place the systems necessary to train, assess, and reward exceptional teaching,” Quick, the provost, and Elizabeth Graddy, vice provost, said in a March memo to the faculty. “In short, let’s make USC the great research university that expects, supports, and truly values teaching and mentoring.”

Clark, at the campus Center for Excellence in Teaching, is helping USC put its money where its mouth is. She said its new model of peer evaluation involves defining teaching excellence and developing training for the faculty, from graduate students who will become professors to full professors.

Peer Review Instead

Peer review will be based on classroom observation and review of course materials, design and assignments. Peer evaluators also will consider professors’ teaching reflection statements and their inclusive practices.

Rewards for high quality teaching will include grants and leaves for teaching development and emphasizing teaching performance in merit, promotion and tenure reviews, Clark said. Most significantly, thus far, the university has introduced continuing appointments for qualifying teaching-intensive professors off the tenure track.

Trisha Tucker, an assistant professor of writing and president of the USC’s Dornsife College of Letters, Arts and Sciences Faculty Council, said different professors have had different reactions to the “culture shift.” But she said she applauded the institution’s ability to resist the “easy shorthand” of teacher ratings in favor of something more meaningful -- albeit more difficult. (USC also has made clear that research and service expectations will not change.)

“It does take work to do this peer review,” she said. “But teaching is important and it takes a lot of time and resources to make that more than just empty words.”

As writing is a feedback-driven process, Tucker said her program already emphasizes pedagogy and peer review. But professors in some other programs will have to adjust, she said.

“For the many faculty who haven’t been trained in this way or hired based on these expectations, it can produce some anxiety,” she said. So an important measure of this new approach’s success is how USC supports people who “initially fall short.”

Clark said the teaching center offers a model for peer review that individual programs will adjust to their own needs over the next year. That kind of faculty involvement in shaping peer review should make for a process that is less "threatening" than representative of an "investment in each other's success," she said.

In the interim, professors’ teaching will be assessed primarily on their own teaching reflections. And while the center avoids using words such as “mandatory” with regarding to training, it is offering a New Faculty Institute, open to all instructors, for 90 minutes monthly over lunch for eight months. Sample topics include active learning, maximizing student motivation and effective, efficient grading practices.

Not Just USC

Philip B. Stark, associate dean of the Division of Mathematical and Physical Sciences and a professor of statistics at the University of California at Berkeley who has studied SETs and argued that evaluations are biased against female instructors in so many ways that adjusting them for that bias is impossible, called the USC news “terrific.”

“Treating student satisfaction and engagement as what they are -- and I do think they matter - - rather than pretending that student evaluations can measure teaching effectiveness is a huge step forward,” he said. "I also think that using student feedback to inform teaching but not to assess teaching is important progress.”

Stark pointed out that the University of Oregon also is on the verge of killing traditional SETs and adopting a Continuous Improvement and Evaluation of Teaching System based on non-numerical feedback. Under the system, student evaluations would still be part of promotion decisions, but they wouldn't reduce instructors to numbers.

Elements of the program already have been piloted. Oregon’s Faculty Senate is due to vote on the program as a whole this week, to be adopted in the fall. The proposed system includes a midterm student experience survey, an anonymous web-based survey to collect non-numerical course feedback to be provided only to the instructor, along with an end-of-term student experience survey. An end-of-term instructor reflection survey also would be used for course improvement and teaching evaluation. Peer review and teaching evaluation frameworks, customizable to academic units, are proposed, too.

“As of Fall 2018, faculty personnel committees, heads, and administrators will stop using numerical ratings from student course evaluations in tenure and promotion reviews, merit reviews, and other personnel matters,” reads the Oregon’s Faculty Senate’s proposal. “If units or committees persist in using these numerical ratings, a statement regarding the problematic nature of those ratings and an explanation for why they are being used despite those problems will be included with the evaluative materials.”

The motion already has administrative support, with Jayanth R. Banavar, provost, soliciting pilot participants on his website, saying, “While student feedback can be an important tool for continual improvement of teaching and learning, there is substantial peer-reviewed evidence that student course evaluations can be biased, particularly against women and faculty of color, and that numerical ratings poorly correlate with teaching effectiveness and learning outcomes.”

More than simply revising problematic evaluation instruments, the page says, Oregon “seeks to develop a holistic new teaching evaluation system that helps the campus community describe, develop, recognize and reward teaching excellence.” The goal is to “increase equity and transparency in teaching evaluation for merit, contract renewal, promotion and tenure while simultaneously providing tools for continual course improvement.”

Craig Vasey, chair of classics, philosophy and religion at the University of Mary Washington and chair of AAUP’s Committee on Teaching, Research and Publications, said the “most pernicious element” of quantitative student evaluations is that the results “get translated into rankings, which then take on a life of their own and don’t really improve the quality of education.”

Review of syllabi and classroom observation by peers are both more “useful means of evaluating,” he said. “And I think asking students how engaged they were in the class -- and especially if they also ask why -- gets “better input from them than the standard questionnaire.”

Ken Ryalls, president of The IDEA Center for learning analytics and a publisher of SETs, told *Inside Higher Ed* earlier this year that not all evaluations are created equal.

“Our advice: Find a good SET that is well designed and low in bias; use the data carefully, watching for patterns over time, adjusting for any proven bias, and ignoring irrelevant data; and use multiple sources of data, such as peer evaluations, administrative evaluations, course artifacts and self-evaluations, along with the student perspective from SETs,” he said via email.