THE CHRONICLE
of Higher Education

Teaching

April 25, 2010

Rating Your Professors: Scholars Test Improved Course Evaluations

By David Glenn

During the next few weeks, hundreds of thousands of college students will fill out course-evaluation forms. On a scale of one to five, they might be asked to rate the instructor's command of the subject material; whether he or she used class time effectively; or whether the exams covered the most important concepts in the course. In most cases they'll also be given space to add anecdotal comments. (On large campuses, it's statistically certain that at least one student will use that space to write a limerick that rhymes "exam" and "scam.")

For students, the act of filling out those forms is sometimes a fleeting, half-conscious moment. But for instructors whose careers can live and die by student evaluations, getting back the forms is an hour of high anxiety. Some people need an extra glass of wine that day.

And many find the concept of evaluations toxic. "They should be outlawed," says D. Larry Crumbley, a professor of accounting at Louisiana State University at Baton Rouge who recently co-edited a book about the topic. "They have destroyed higher education." Mr. Crumbley believes the forms lead inexorably to grade inflation and the dumbing down of the curriculum.

Outlawing the forms seems unlikely. The tide, in fact, seems to be moving in the opposite direction. Last year, Texas enacted a law that will require every public college to post each faculty member's student-evaluation scores on a public Web site.

So can the evaluations be improved? Various scholars and administrators have tried. Here are four efforts to build a better mousetrap:

Custom Questions

The IDEA Center, an education research group based at Kansas State University, has been spreading its particular course-evaluation gospel since 1975. The central innovation of the IDEA system is that departments can tailor their evaluation forms to emphasize

whichever learning objectives are most important in their discipline. For an anatomy class, the most important objective might be gaining factual knowledge. For a literature course, it might be analytic writing.

Departments can also customize the forms further by adding questions specific to a particular local course. A mandatory information-literacy course at Saint Francis University, in Pennsylvania, has customized its IDEA forms with questions about whether enough time was spent on citation skills, online resources, and critically evaluating information sources.

"When a particular department devises its own set of questions, that can eventually give them a set of data that they can assess over time," says Theresa L. Wilson, an instructional-technology specialist at Saint Francis who helps oversee the system.

The IDEA system is attractive, Ms. Wilson says, because it offers both locally customizable features and the ability to compare certain scores with a large, nationally normed database. (Roughly 350 colleges use the IDEA Center's system, though in some cases only a single department or academic unit participates.) When the IDEA Center analyzes a college's data, it provides both "raw" scores for each course and also scores that are adjusted to account for factors outside the instructor's control. For example, in a large general-education class where many of the students say they were not motivated to learn the material, course-evaluation scores will tend to be lower than in a small class that is composed entirely of majors. The IDEA system's adjusted scores try to correct for such structural biases.

Peter Skoner, associate provost of Saint Francis, agrees that the IDEA system has been valuable, but he says that its forms are only a small part of the university's faculty-evaluation system. "Department chairs visit every class each year," he says. "We have annual reviews that faculty members complete with their chairs. We have self-reflection. We have standardized exam scores. So there are many ways to assess teaching and learning, and these IDEA forms are only a piece of the puzzle."

Student Self-Assessment

More than a decade ago, Elaine Seymour, who was then director of ethnography and evaluation research at the University of Colorado at Boulder, was assisting with a National Science Foundation project to improve the quality of science instruction at the college level. She found that many instructors were reluctant to try new teaching techniques because they feared their course-evaluation ratings

might decline.

Ms. Seymour and her colleagues thought that was a sad dynamic. So they did an end-run around the problem by developing a new evaluation instrument designed to capture students' own perceptions of how much they learned in a course—and, more importantly, which course elements helped them learn.

The survey instrument, known as SALG, for Student Assessment of their Learning Gains, is now used by instructors across the country. The project's Web site contains more than 900 templates, mostly for courses in the sciences.

Like the IDEA Center's model, the SALG allows instructors to customize questionnaires to match the objectives of a particular course. But the forms must also include a minimal number of baseline elements, which allow data to be compared across institutions and over time.

"One of our biggest challenges," says Robert D. Mathieu, a professor of astronomy at the University of Wisconsin at Madison and the principal investigator on the SALG project, "is that we have instructors and departments that want to use the SALG, but at the same time are also required by their institutions to use the classic types of course evaluation. So you end up with a situation where students are double-hit, and there's a certain amount of survey fatigue."

Mr. Mathieu hopes to soon test the use of SALG across an entire college. (There may be a pilot next year at Santa Clara University.) One of the program's latest innovations is a feature that analyzes the texts of students' survey comments, indicating which words are used most frequently and allowing instructors to see patterns that they might otherwise miss.

"If you just flip through students' comments," Mr. Mathieu says, "you tend to give highest weight to the extrema"—that is, the outliers who gush with praise or rip your teaching to shreds. "So the ability to do some quantitative analysis of these comments really allows you to take a more nuanced and effective look at what these students are really saying."

Quality Teaching

The Teaching and Learning Quality survey has been developed during the last three years by Theodore W. Frick, an associate professor of education at Indiana University at Bloomington. Unlike the IDEA Center's model, this questionnaire is meant to be one-size-fits-all.

The project began when Mr. Frick was asked to serve on a teaching-awards committee. The committee looked at nominees' course-evaluation data, but Mr. Frick was deeply skeptical. "I looked at these forms and said, Gee, does this have anything to do with student learning?" he says. "Like many faculty members, I was pretty jaundiced about the entire concept of course evaluations. I just thought it was a smiles test and a test of popularity."

He started to read the scholarly literature on course evaluations and reluctantly concluded that there is at least a weak relationship between students' global ratings of a course and certain measures of their learning. On the other hand, Mr. Frick thought his campus's evaluation forms could be improved if they included items about teaching practices that are known to improve student learning. So he drafted a form that asked about how effectively the instructors activated the students' previous knowledge, demonstrated skills and concepts, and offered applications. For instance, students are asked whether they agree with statements like, "In this course, I was able to reflect on, discuss with others, and defend what I learned."

In a pilot study that included 12 courses in business, philosophy, kinesiology, and several other fields, Mr. Frick and his colleagues found that his new course-evaluation form was strongly correlated with both students' and instructors' own measures of how well the students had mastered each course's learning goals.

Mr. Frick wants his instrument to be used to help instructors improve, not as a high-stakes measure for hiring and firing instructors. "When course evaluations are a major part of tenure and promotion, I think over time that gives faculty members an incentive to design the evaluations so that they'll get good ratings, whether or not they have anything to do with student learning," says Mr. Frick, who will discuss his project this week at the annual meeting of the American Educational Research Association.

Minimal Bias

Two years ago, when it became clear that the Texas legislature was likely to require public online disclosure of faculty members' course-evaluation scores, officials at the University of North Texas got very anxious. The university had no standard course-evaluation system, and the forms varied enormously across various academic units. That meant that the evaluation scores that would soon become public were nowhere near apples-to-apples comparisons. Administrators and faculty members foresaw a lot of unhappiness—and maybe even lawsuits.

The provost asked a team of experts in psychometrics and human

resources to come up with a campuswide course-evaluation tool that would minimize any distortions or biases. The committee convened focus groups with more than 300 faculty members and 80 students. "For example, we would look at some proposed items and ask the focus groups whether students were actually in a position to measure those items," says Paula Iaeger, a graduate student in education who served on the committee. "Faculty members and students both told us, for example, that students should not be asked whether the textbook was the best possible for the course, because students can't know that."

Ms. Iaeger says her committee took pains to include adjunct faculty members in those focus groups. "If we were only listening to our tenure-track faculty, we would have gotten an incomplete picture," she says.

The new North Texas instrument that came from these efforts tries to correct for biases that are beyond an instructor's control. The questionnaire asks students, for example, whether the classroom had an appropriate size and layout for the course. If students were unhappy with the classroom, and if it appears that their unhappiness inappropriately colored their evaluations of the instructor, the system can adjust the instructor's scores accordingly.

"And we can also use that data in other ways," Ms. Iaeger says. "Once we accumulate enough data about which classrooms the students dislike, we can try not to assign novice teachers to those classrooms, so the students don't get a double-whammy."

Don't Ask the Cars

None of these innovations impresses Mr. Crumbley, the Louisiana State skeptic. No matter how sophisticated a survey instrument might be, he says, students are too likely to answer in haphazard or dishonest ways.

"Students are the inventory," Mr. Crumbley says. "The real stakeholders in higher education are employers, society, the people who hire our graduates. But what we do is ask the inventory if a professor is good or bad. At General Motors," he says, "you don't ask the cars which factory workers are good at their jobs. You check the cars for defects, you ask the drivers, and that's how you know how the workers are doing."

Few critics of course evaluations are willing to go that far. William H. Pallett, president of the IDEA Center, says that when course rating surveys are well-designed and instructors make clear that they care about them, students will answer honestly and

thoughtfully.

"Student ratings aren't the be-all and end-all," Mr. Pallett says. "But they can inform instructors about things they can do to improve their learning. And students usually do correctly identify the tasks and learning objectives that were most important to a course. If they were just filling out these forms completely nonchalantly," he says, "we wouldn't see that pattern."

Another who finds value in evaluations is Ken Bain, vice provost of instruction at Montclair State University and author of *What the Best College Teachers Do* (Harvard University Press). Almost everyone agrees that course evaluations by themselves are inadequate, he says. But both faculty members and administrators (perhaps for different reasons) are too hesitant to add less-quantitative assessments of teaching.

In Mr. Bain's view, student evaluations should be just one of several tools colleges use to assess teaching. Peers should regularly visit one another's classrooms, he argues. And professors should develop "teaching portfolios" that demonstrate their ability to do the kinds of instruction that are most important in their particular disciplines.

"It's kind of ironic that we grab onto something that seems fixed and fast and absolute, rather than something that seems a little bit messy," he says. "Making decisions about the ability of someone to cultivate someone else's learning is inherently a messy process. It can't be reduced to a formula."