

What Are Students Rating When They Rate Instructors?

IHE [insidehighered.com/news/2017/05/10/study-student-ratings-instructors-dependent-discipline-quantitative-fields-are-most](https://www.insidehighered.com/news/2017/05/10/study-student-ratings-instructors-dependent-discipline-quantitative-fields-are-most)

Student evaluations of teaching, or SET, aren't short on critics. Many professors and other experts say they're unreliable -- they [may hurt female and minority professors](#), for example. [One recent metastudy](#) also suggested that past analyses linking student achievement to strong evaluation scores are flawed, a mere "artifact of small-sample-sized studies and publication bias."

Now one of the authors of that metastudy is back for more, with a new analysis suggesting that SET ratings vary by course subject, with professors of math-related fields bearing the brunt of the effect.

"Professors teaching quantitative courses are far more likely not to receive tenure, promotion and/or merit pay when their performance is evaluated against common standards," reads the study, co-written by SET skeptic Bob Uttl, professor of psychology at Mount Royal University in Canada.

Uttl's co-author for the new study is Dylan Smibert, a Ph.D. candidate in organizational and industrial psychology at Saint Mary's University, also in Canada. Attempting to examine the influence of course type on how students rated their instructors and related personnel decisions, Uttl and Smibert analyzed 14,872 publicly available class evaluations from New York University. Those summary evaluations were based on some 325,538 individual student ratings. Classes considered were those in math and statistics and English, history, and psychology, which the authors use as stand-ins for quantitative and nonquantitative fields, respectively.

They found that class subject is strongly associated with SET ratings and has a substantial impact on professors being labeled satisfactory vs. unsatisfactory and excellent vs. nonexcellent. The impact varies substantially depending on criteria on which the evaluations are based, the authors say, but the article not-so-subtly suggests that teaching quantitative courses is "hazardous" to one's career. ("[Student Evaluations of Teaching: Teaching Quantitative Courses Can Be Hazardous to One's Career](#)" is available online in *PeerJ*.)

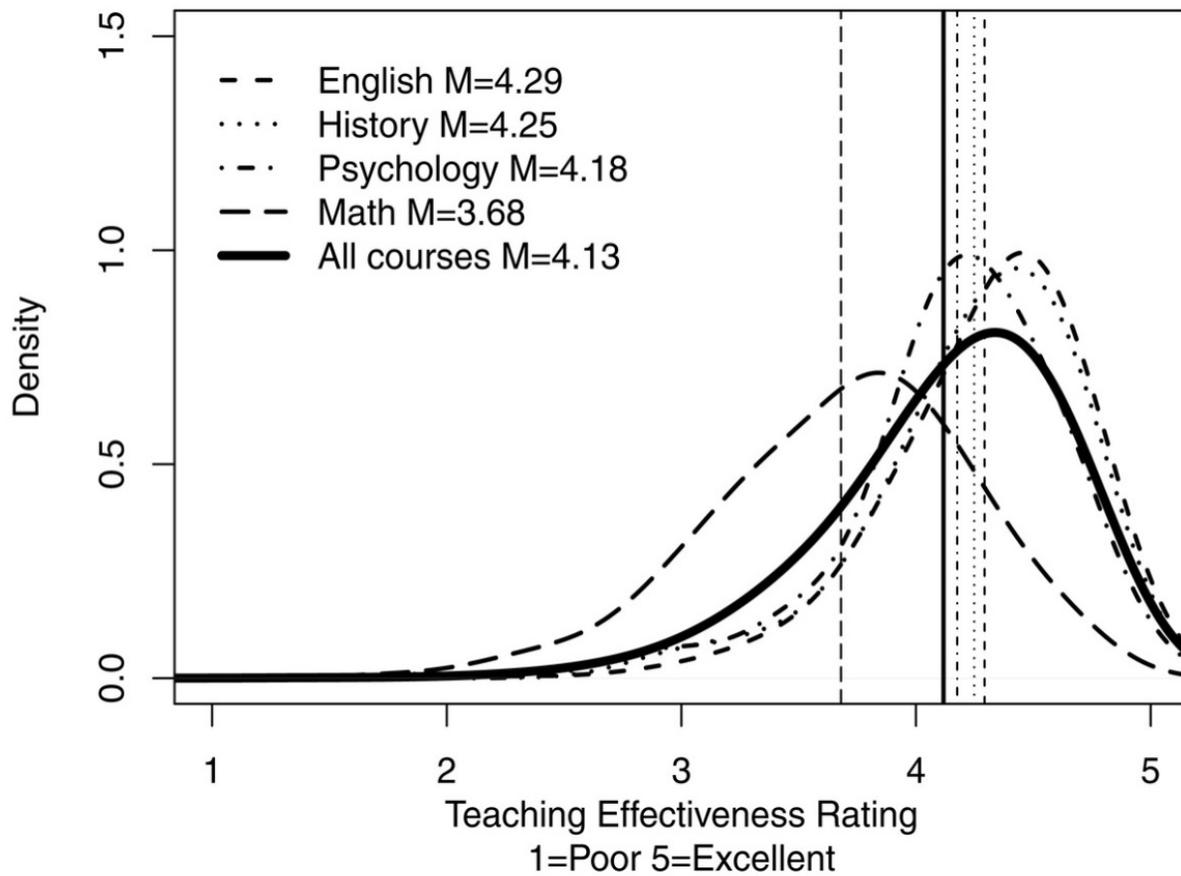
Professors teaching quantitative vs. nonquantitative courses are far more likely to fail norm-referenced cutoffs, the study says. They're 1.88 times more likely not to meet the overall mean standard and 1.27 times more likely to fail the "excellent" standard, 3.17 times more likely to fail the "very good" standard, and 6.02 times more likely to fail the "good" standard.

Here are some of the questions asked in the NYU's evaluations, along with mean ratings and standard deviations in the courses studied, on a scale of 1 (poor) to 5 (excellent):

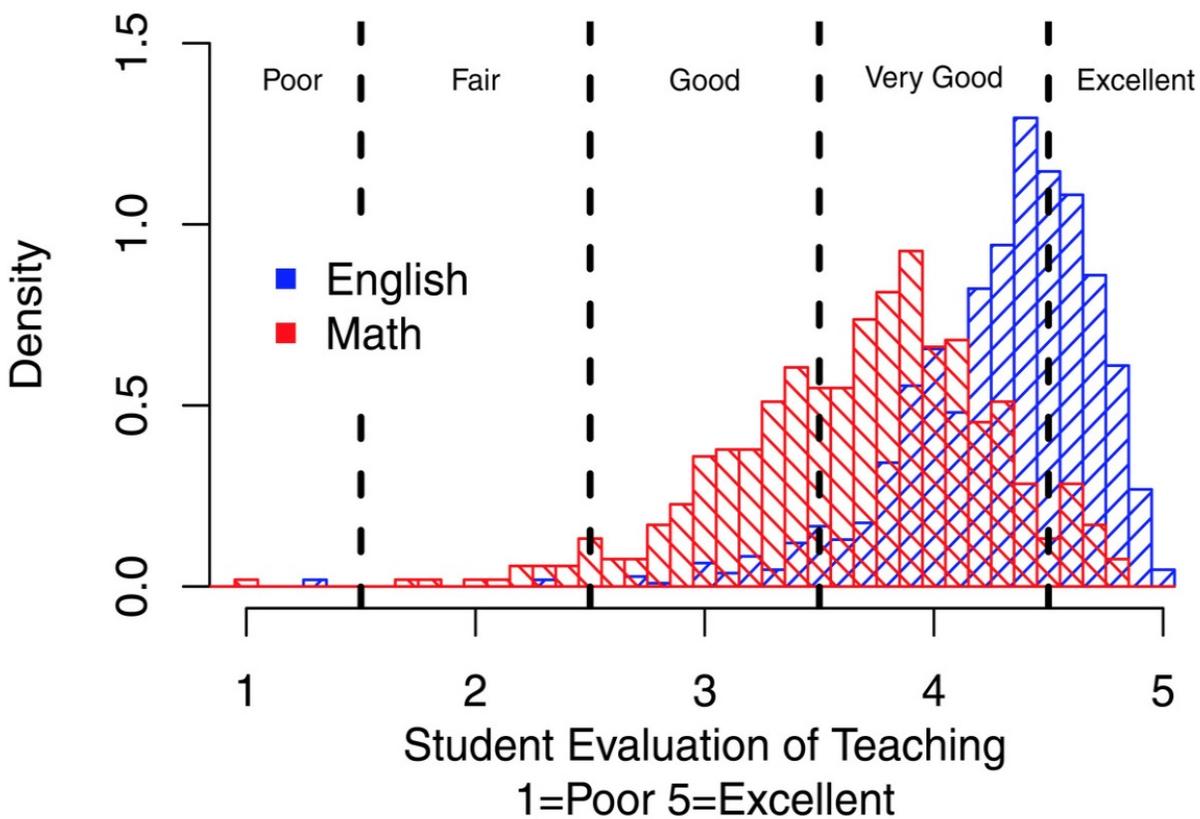
SET questions, mean ratings, and standard deviations.

| Question | <i>M</i> | <i>SD</i> |
|---|----------|-----------|
| 1. How would you rate the instructor overall? | 4.37 | 0.55 |
| 2. How informative were the classes? | 4.26 | 0.52 |
| 3. How well organized were the classes? | 4.19 | 0.55 |
| 4. How fair was grading? | 4.14 | 0.55 |
| 5. How would you rate this course overall? | 4.09 | 0.57 |
| 6. How clear were the objectives of this course? | 4.12 | 0.52 |
| 7. How well were these objectives achieved? | 4.10 | 0.53 |
| 8. How interesting was the course? | 4.01 | 0.63 |
| 9. To what extent were your own expectations met? | 3.90 | 0.58 |
| Mean overall rating (across all items) | 4.13 | 0.50 |

And here are distributions for those ratings based on the course subject. Notice that math is significantly to the left (and lower than) the vertical mean line:



The difference is starker when comparing only math and English, below. While 71 percent of English courses pass the overall mean as the standard, just 21 percent of math courses do. "The vast majority of math courses (79 percent) earn their professors an 'unsatisfactory' label in this scenario," the study says.



Uttl and Smibert note that previous research has found that irrelevant factors related to teaching effectiveness, such as class size or discipline, don't make a difference in student ratings (and that student ratings are therefore reliable markers of teaching quality). But, the authors say, such research has been "plagued" by methodical shortcomings -- namely that it doesn't account for "severe ceiling effects," or a large proportion of students giving professors the highest possible ratings. Other problems include that scaled ratings are often used to making "binary" decisions between good or bad teaching -- so a professor who missed a department's mean SET rating by the narrowest of margins, for example, might be terminated while a colleague who scored marginally higher would be safe. Studies also often look at individual student ratings of teachers and not overall class summaries, the paper says.

"Our results show that math classes received much lower average class summary ratings than English, history, psychology or even all other classes combined, replicating previous findings showing that quantitative vs. nonquantitative classes receive lower SET ratings," the study reads. "More importantly, the distributions of SET ratings for quantitative vs. nonquantitative courses are substantially different." Whereas the SET distributions for nonquantitative courses show a typical negative skew and high mean ratings, it says, distributions for quantitative courses are less skewed, nearly normal, and have substantially lower ratings.

The passing rates for various common standards for effective teaching are substantially lower for professors teaching quantitative vs. nonquantitative courses, the authors add. "Clearly, professors who teach quantitative vs. nonquantitative classes are not only likely to receive lower SETs, but they are also at a substantially higher risk of being labeled 'unsatisfactory' in teaching, and thus, more likely to be fired, not reappointed, not promoted, not tenured and denied merit pay."

Such outcomes will be most apparent -- or unfair -- in departments in which some professors teach quantitative courses and others teach nonquantitative courses, such as psychology or sociology, suggests the study.

Is the answer, then, to do away with SETs? At the very least, Uttl and Smibert say, "fairness requires that we evaluate a professor teaching a particular subject against other professors teaching the same subject rather than against some common standard."

Used this way, they add, "SET ratings can at least tell us where a professor stands within the distribution of other professors teaching the same subjects, regardless of what SETs actually measure."

There are, of course, those who defend student ratings of instruction as important gauges of students' classroom experiences. IDEA, a popular SET vendor, for example, offers relatively robust services in that it controls for things such as class size, provides comparative scores among instructors and offers specified feedback on ways to improve one's teaching.

Ken Ryalls, IDEA's president, said researchers have long known that certain courses and disciplines result in lower ratings for instructors for a variety of reasons. At the same time, he said, the discussion shouldn't be "whether or not to use student feedback in tenure and promotion decisions, but rather how to use the data." That's because "the student voice matters."

Ryalls said the ultimate challenge, then, is that "we use all student ratings cautiously when considering them in tenure and promotion decisions. Colleges should adjust the weight they give to student ratings of instruction depending on factors outside of the instructor's control, including discipline."

Uttl said Tuesday that the inescapable issue is still "what it is that SETs measure. Do they measure student satisfaction or do they measure professors' teaching effectiveness?" Consulting *Merriam-Webster*, he said satisfaction is "a fulfillment of a need or want" or "a happy or pleased feeling because of something that you did or something that happened to you" -- something very different from effective teaching.

Another problem, Uttl added, is that “there is no agreement as to what is effective teaching.” Strong teaching will vary with discipline, he said, and “we have no universal agreed-upon standard measure of effective teaching.”

If SETs measure satisfaction rather than professors’ teaching effectiveness, what should colleges and universities do? Uttl said they shouldn’t be used for high-stakes personnel decisions or to judge the quality of one’s teaching. Policy makers and institutions also need to decide what’s more important, he said -- students’ satisfaction or their academic achievement.