

# What Does the Research Say About Students Evaluating Their Teachers?

By Pieter Spooren, Bert Brockx, and Dimitri Mortelmans

In this important article in *Review of Educational Research*, Pieter Spooren, Bert Brockx, and Dimitri Mortelmans (University of Antwerp) report on their meta-analysis of recent research on student evaluations of college instructors. In the higher-education world, student surveys are the most common measure of teaching quality – in some cases, the only measure. Students’ opinions are easy to gather, and students can be seen as the most important stakeholders to ask – as Peter Seldin puts it, “the opinions of those who eat the dinner should be considered if we want to know how it tastes.” (1993)

Student evaluation of teaching can potentially serve three purposes, say Spooren, Brockx, and Mortelmans:

- Feedback to improve the quality of teaching;
- Evaluative information on instructors for tenure and promotion decisions;
- Quality assurance and evidence of institutional accountability to outside stakeholders.

There’s always been tension between the formative and summative aspects of student evaluations – improving teaching versus decisions on instructors’ professional careers. The key question in either case, but especially the latter, is the validity and reliability of students’ opinions of their teachers. So are student surveys valid and reliable?

Answering that question is complicated, say Spooren, Brockx, and Mortelmans, because a number of factors affect the quality of student surveys:

- *How they are seen* – Evaluations have been called “happy forms,” “personality contests,” and measures of “customer satisfaction.” If students don’t have a sense that their feedback will be taken seriously, they may not be careful and thoughtful as they fill out the forms. There can also be a “halo effect,” with students who give high ratings in one area giving high ratings in others.
- *Survey quality* – Some student questionnaires are poorly worded and haven’t been tested for their psychometric properties.
- *Common conceptual framework* – If students and instructors don’t have a shared understanding of what constitutes effective teaching, the impact of survey results can be suboptimal.
- *Anonymity* - When students are able to submit their opinions anonymously (which is almost universally the case), does that lead to depersonalizing the relationship between instructors and students? In anonymous surveys, students’ opinions “disappear without a trace” into the aggregate data, say the authors, and there’s rarely a space for “discussing, explaining, or negotiating the results with students.”

- *Electronic versus paper questionnaires* – Response rates are much lower when students fill out surveys online (29% versus 70% in one study), but results are comparable and online surveys tend to generate longer and more thoughtful comments.
- *Interpretation and use* – Making meaning of survey results is more difficult than it looks, say the authors, and there's the risk of inappropriate use by administrators if guidelines and training aren't in place.
- *Faculty reactions* – Most instructors care about their students' opinions and are anxious when receiving them. Many faculty members are not aware of the positive research on student evaluations, and the less they know, the more they believe a number of persistent myths. Most faculty members don't find students' comments very helpful and/or ignore the feedback.
- *Gaming the system* – Some instructors focus on trying to improve their student-evaluation scores rather than using students' feedback to improve their teaching – for example, grading more leniently or giving less-rigorous assignments.

What were the findings of this meta-analysis? First, that the data from student surveys tend to be highly stable over time, indicating that students in different courses in different years are identifying important common themes about each teacher's work.

Second, there is a positive correlation between student survey results and other measures of teaching quality – for example, student learning outcomes, self-ratings, and alumni ratings. The most rigorous analysis would compare student evaluations with class test averages, common tests in multiple-section courses, pre- and post-tests, monitoring achievement in future classes, and using standardized assessments.

Third, instructors tend not to use students' feedback to reflect on and improve their teaching. This explains why, over time, student surveys have not led to improvements in teaching. Student survey results are much more likely to have a positive impact when instructors self-assess and look at the data with a colleague, administrator, or expert observer. (However, teachers with lower ratings tend to overestimate what students will say, and teachers with higher ratings tend to underestimate.)

Finally, the results of official student surveys correlate quite well with unofficial RateMyProfessor-type websites, but the authors say that “student evaluations from these websites should be interpreted with great caution.” Why? Because of selection bias (students who use them tend to have very strong positive or negative opinions) and unscientific data, including opinions on teachers' “hotness” and “sexiness.” The authors also caution against using any student survey data as a sole criterion for evaluating instructors.

“On the Validity of Student Evaluation of Teaching: The State of the Art” by Pieter Spooren, Bert Brockx, and Dimitri Mortelmans in *Review of Educational Research*, December 2013 (Vol. 83, #4, p. 598-642)