Student Ratings Myths vs. Research

Michael Theall, PhD

Student ratings of instruction are hotly debated on many college campuses. Unfortunately these debates are often uninformed by the extensive research on this topic. Marsh’s often-cited review of the research on student ratings shows that student ratings data are (a) multidimensional; (b) reliable and stable; (c) primarily a function of the instructor who teaches the course; (d) relatively valid against a variety of indicators of effective teaching; (e) relatively unaffected by a variety of variables hypothesized as potential biases; and (f) seen to be useful by faculty, students, and administrators. The researchers who have synthesized all the major studies of ratings have reached the same conclusions as Marsh. But even when the ratings data are technically rigorous, one of the major problems is day-to-day practice: student ratings are often misinterpreted, misused, and not accompanied by other information that allows users to make sound decisions. As a result, there is a great deal of suspicion, anxiety, and even hostility toward ratings. Several questions are commonly raised with respect to student ratings. Current research provides answers to many of these questions.

1. Are students qualified to rate their instructors and the instruction they receive?

Generally speaking, the answer is yes. Students can report the frequencies of teacher behaviors, the amount of work required, how much they feel they have learned, and the difficulty of the material. They can answer questions about the quality of lectures, the value of readings and assignments, the clarity of the instructor’s explanations, the instructor’s availability and helpfulness, and many other aspects of the teaching and learning process. No one else is as qualified to report what transpires during the semester, simply because no one else is there for the entire semester. Students are certainly qualified to express their satisfaction or dissatisfaction with the experience. They have a right to express their opinions in any case, and no one else can report the extent to which the experience was useful, productive, informative, satisfying, or worthwhile. Although opinions on these matters are not direct measures of the performance of the teacher or the content learned, they are legitimate indicators of student satisfaction; there is a substantial research base linking this satisfaction to effective teaching and learning. But students are not necessarily qualified to report on all issues. For example, beginning students cannot accurately rate the instructor’s knowledge of the subject. A colleague’s rating is more appropriate for this purpose. Likewise, peers are better qualified to judge content currency, curricular match, course design, or assessment methods. Both students and peers are in unique positions to provide enlightening perspectives. For effective evaluation, remember to use multiple sources of data and ask questions that respondents can legitimately answer.

2. Are ratings based solely on popularity?

There is no basis for this argument and no research to substantiate it. When this topic arises, the term popular is never defined. Rather, it is left to imply that learning should somehow be
unpleasant, and the popularity statement is usually accompanied by an anecdote suggesting “The best teacher I ever had was the one I hated most.” The assumption that popularity somehow means a lack of substance, knowledge, or challenge is entirely without merit. In fact, several studies show students learn more in courses in which teachers demonstrate interest/concern for the students and their learning. Of course these teachers also receive higher ratings.

3. Are ratings related to learning?

The most acceptable criterion for good teaching is student learning. There are consistently high correlations between student ratings of the “amount learned” in a course and students’ overall ratings of the teacher and the course. Even more telling are the studies in multisection courses that employed a common final exam. In general, student ratings were the highest for instructors whose students performed best on the exams. These studies are the strongest evidence for the validity of student ratings because they connect ratings with learning.

4. Are ratings affected by situational variables?

The research says that ratings are robust and not greatly affected by situational variables. But we must keep in mind that generalizations are not absolute statements. There will always be some variations. For example, we know that required large-enrollment, out-of-major courses in the physical sciences get lower average ratings than elective, upper-level major courses in virtually all other disciplines. Does this mean that teaching quality varies? Not necessarily. What it does show is that effective teaching and learning may be harder to achieve under certain sets of conditions. There is a critical principle for evaluation practice embedded here. To be fair, comparisons of faculty teaching performance based on ratings should use sufficient amounts of data from similar situations. It would be grossly unfair to compare the ratings of an experienced professor teaching a graduate seminar of ten students to the one-time ratings of a new instructor teaching an entry-level required course with an enrollment of 300.

5. Do students rate teachers on the basis of expected (or received) grades?

This is currently the most contentious question in ratings research. There is consistent evidence of a relationship between grades and ratings: a modest correlation of about .20. The multisection validity studies (mentioned in question 3) provide the most solid evidence that ratings reflect learning (a correlation of about .43). These findings lead to the conclusion reached by most researchers: There should be a relationship between ratings and grades because effective teaching leads to learning that leads to student achievement and satisfaction. Ratings simply reflect this sequence.

6. Can students make accurate judgments while still involved in their schooling?

Some argue that students cannot discern real quality until years after leaving the classroom. There is no research proving this statement. However, several studies compare in-class ratings to ratings by the same students the next semester, the next year, immediately after graduation, and several years later. All these studies report the same results: Although students may realize later that a particular subject was more or less important that they thought, student opinions about
teachers change very little over time. Teachers rated highly in class are rated highly later on, and those with poor ratings in class continue to get poor ratings later on. This question is connected to the larger technical matter of overall reliability of ratings. The research indicates that ratings are very reliable. Whether reliability is measured within classes, across classes, over time, or in other ways, student ratings are remarkably consistent.

Editor’s note: Michael Theall has twenty-six years of experience as a faculty member and as a professional in instructional design, development, and evaluation. He has founded faculty centers for teaching, learning, and evaluation at three universities: the University of Illinois, the University of Alabama, and Youngstown State University (OH). Theall and colleague Jennifer Franklin recently received a career achievement award from the American Education Research Association (AERA) for their work in faculty evaluation and development. They are authors of “The Student Ratings Debate,” a monograph for New Directions for Institutional Research (2001), among numerous other publications on evaluating teaching. Theall penned this article for BYU’s Focus on Faculty Newsletter (vol. 10, no. 3, p.2) at the request of the editor.

Notes


Other Works Cited and a Bibliography of Recent Work


