Random Thoughts . . .

STUDENT RATINGS OF TEACHING: MYTHS, FACTS, AND GOOD PRACTICES

RICHARD M. FELDER
North Carolina State University
REBECCA BRENT
Education Designs, Inc.

Myths about student ratings of teaching abound on every campus, usually accompanied by confident assurances that ratings are just popularity contests that reward entertainers and penalize the best teachers. (Remarkably, the second group always seems to include the ones doing the assuring.) Some years ago we surveyed the myths and summarized the extensive research that showed most of them to be wrong.[1] Now it's 15 years later and a lot more research has been done, with similar outcomes. Unfortunately the myths are still alive and well, so here is the 2008 version of what "everyone knows" about student ratings and how much of that wisdom is supported by research.

* * *

Myth. *Student ratings are not valid assessments of teaching quality.*

Fact. False. Thousands of research studies have shown that student ratings correlate positively with every other measure of teaching effectiveness, including alumni ratings, peer ratings, administrator ratings, measures of learning (*e.g.*, standardized tests, common exams in multi-section courses, and ratings of student portfolios), and student motivation.[2] The magnitude of the observed correlations varies considerably across individual studies and a few studies report contradictory results, but the weight of the evidence is clear. If students consistently say someone's teaching is good or bad, they're almost certainly right.

Myth. *The highest ratings go to the easiest courses.*

Fact. False. Up to a point, courses rated as more difficult on average get higher ratings than easier courses, with ratings only beginning to fall when courses reach levels of difficulty beyond the backgrounds of most enrolled students. In a recent study of 1,045 engineering, science, and humanities courses at two universities, Dee[3] found that student perceptions of course workload were not significantly different for courses in the top and bottom quartiles of student ratings, with the marginally higher workload rating going to the courses in the top quartile.

Myth. *Bad teachers who are easy graders get higher evaluations than good teachers who are strict graders.*

Fact. False. Individual instructors who give high grades relative to local averages may get higher ratings than they would if their grades were lower,[4] but no studies have turned up ineffective teachers who got high ratings just by giving high grades.[5] However, the possibility that it *could* happen supports the common recommendation to use multiple sources of assessment data.

Myth. *They may not like me now because I'm rigorous and maintain high standards, but in a few years they'll appreciate how good a teacher I was.*

Fact. Generally false—it happens sometimes, but not often.[1, 2] Alumni ratings correlate significantly with student ratings given previously to the same instructor. If your students think you're a great teacher now, most will still remember you fondly in the future, and if they think you're lousy, don't

Richard M. Felder *is Hoechst Celanese Professor Emeritus of Chemical Engineering at North Carolina State University. He is co-author of* Elementary Principles of Chemical Processes *(Wiley, 2005) and numerous articles on chemical process engineering and engineering and science education, and regularly presents workshops on effective college teaching at campuses and conferences around the world. Many of his publications can be seen at* <www.ncsu.edu/felder-public>.

Rebecca Brent *is an education consultant specializing in faculty development for effective university teaching, classroom and computer-based simulations in teacher education, and K-12 staff development in language arts and classroom management. She codirects the ASEE National Effective Teaching Institute and has published articles on a variety of topics including writing in undergraduate courses, cooperative learning, public school reform, and effective university teaching.*

expect to start getting holiday greetings from them in five or 10 years.

* * *

Not all common beliefs about evaluations are wrong, of course. It's true that teachers who are enthusiastic and caring tend to get better ratings than those who are reserved and distant, but so what? Enthusiasm and caring of instructors also correlate with motivation and learning of their students,[2] suggesting that the higher ratings are probably deserved. It's also true that other things being equal, elective courses tend to get higher ratings than required courses, upper-level courses tend to get higher ratings than lower-level courses, small and moderately sized classes get higher ratings than very large classes, student ratings in engineering and the sciences are lower than ratings in other fields, and female instructors in engineering and the physical sciences get lower ratings than male instructors.[2] On average these effects are small, but they exist and should be taken into account when ratings are used to make decisions about such things as reappointment, tenure, promotion, and merit raises.

In short, student evaluations have high levels of reliability and validity and should always be part of the process used to evaluate teaching. There are some aspects of a course that students are in no position to evaluate, however, including whether the course learning objectives are appropriate, the content is current with the state of the field, and the course adequately prepares the students for subsequent courses in the curriculum. Those things can only be evaluated by knowledgeable peers. Student ratings should therefore not be the sole source of teaching assessment data but should be supplemented with peer ratings and other measures of teaching effectiveness.[6] If different sources agree, as they usually will, it's a good indication that the overall assessment is a fair one; if they disagree, it's a red flag, and an effort should be made to find out what's going on.

Since student ratings will undoubtedly remain central to teaching evaluation (as they should), everything possible should be done to make them as effective as possible. The following recommendations—most of which are drawn from the papers cited in Reference 2—address that goal.

- *Use a rating form that has been developed with the assistance of someone knowledgeable about educational measurements.* There is a science to survey construction in general and educational rating instrument construction in particular. Either use a form that has been developed and validated elsewhere, such as the IDEA Student Ratings of Instruction system (<*http://www.idea.ksu.edu/*>) or TCETools (<*http://tcetools.com*>), or work with an education specialist on your campus or an external consultant.
- *Don't trust ratings collected from fewer than 10 students or less than 2/3 of a class, and don't make personnel decisions based on ratings from a single semester.*
- *Use a few global or summary items with Likert-scale (1–5) ratings for summative evaluation* (evaluation used to help inform personnel decisions), *and a longer list of more specific items for formative evaluation* (diagnostic evaluation used to help instructors improve their teaching). Global items correlate more strongly with student learning than more specific items do, and you'll get a higher rate of return if there are fewer questions.
- *When evaluating ratings, remember that they may be slightly affected by factors other than the quality of instruction, such as the nature and level of the course, the class size, and the gender of the instructor.* The IDEA system and TCETools include provisions for taking these factors into account.
- *Try to persuade students that their ratings will be considered carefully and may have an impact on faculty personnel decisions and decisions about teaching assignments.* If you can make this case convincingly, most students will take the ratings seriously and you should get a good rate of return. If you can't make the case, there is no reason the students *should* take the ratings seriously and you should not be surprised if they don't.

REFERENCES

1. Felder, R.M., "What Do They Know, Anyway?" *Chem. Eng. Ed.*, **26**(3), 134-135 (1992), <http://www.ncsu.edu/felder-public/Columns/Eval.html>
2. (a) Cashin, W.E., "Student Ratings of Teaching: Recommendations for Use," IDEA Paper No. 22, Center for Faculty Evaluation and Development, Kansas State University, January 1990, <http://www.idea.ksu.edu/papers/Idea_Paper_22.pdf>; (b) Ory, J.C., "Faculty Thoughts and Concerns About Student Ratings," *New Directions for Teaching and Learning*, **87**, 3-15 (Fall 2001); (c) McKeachie, W.J., "Student Ratings: Their Validity of Use," *American Psychologist*, **52**(11), 1218-1225 (1997); (d) McKeachie, W.J., *Teaching Tips*, 11th Ed., Boston, Houghton Mifflin Co., 326-328 (2002)
3. Dee, K.C., "Student Perceptions of High Course Workloads Are Not Associated With Poor Student Evaluations of Instructor Performance," *J. Eng. Ed.*, **96**(1), 69–78 (2007)
4. Greenwald, A.G., and G.M. Gillmore, "Grading Leniency is a Removable Contaminant of Student Ratings," *American Psychologist*, **52**(11), 1209-1217 (1997)
5. Abrami, P.C., W.J. Dickens, R.P. Perry, and L. Leventhal, "Do Teacher Standards for Assigning Grades Affect Student Evaluations of Instruction?" *J. Educational Psychology*, **72**, 107–118 (1980)
6. Felder, R.M., and R. Brent, "How to Evaluate Teaching," *Chem. Eng. Ed.*, **38**(3), 200-202 (2004), <http://www.ncsu.edu/felder/public/Columns/Teacheval.pdf>. See also "A Protocol for Peer Review of Teaching," (<http://www.ncsu.edu/felder-public/Papers/ASEE04(Peer-Review).pdf>) and "If You've Got It, Flaunt It: Uses and Abuses of Teaching Portfolios" (<http://www.ncsu.edu/felder-public/Columns/Portfolios.html>) ❐

All of the *Random Thoughts* columns are now available on the World Wide Web at http://www.ncsu.edu/effective_teaching and at http://che.ufl.edu/~cee/