

Prepared for the American Psychologist

Published in *American Psychologist*, 53 (1998), 1223-1224.

Are Student Ratings of Instruction Useful?

J. Scott Armstrong
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104

E-mail: armstrong@wharton.upenn.edu
phone 215-898-5087
fax 215-898-2534

June 1998

Despite the lead article's title "Validity Concerns and Usefulness of Student Ratings of Instruction" (Greenwald 1997) in the *American Psychologist*'s special section on teacher ratings, the papers did not provide direct evidence on "usefulness." There is no evidence that the use of teacher ratings improves learning in the long run. The papers do not show that the effects would improve the allocation of effort between teaching and research, or that the quality of the educational experience will be better, or that students and faculty will be happier. Given the evidence to date, the case for student ratings is weak. I raise some questions about usefulness, with a particular emphasis on the ratings' effects on learning.

Are teacher ratings related to learning? The correspondence of ratings to learning should be very close, otherwise, it is not clear what will change. When I was involved with worker incentive standards, the prevailing wisdom was that even modest departures from the desired measure created dysfunctional outcomes. Perhaps the most important finding from the special section is that teacher ratings and learning are not closely related.

Furthermore, the ratings/learning relationship seems to be based heavily on studies involving rote learning. The studies provide no breakdown of evidence for studies where skill

development is the goal. Consider, for example, a study by Attiyeh and Lumsden (1972), that examined the relationship between teacher ratings and common final exams on economics. This study covered over 30,000 students in the U.K. and it attempted to assess students' skills in applying concepts, rather than to assess content learning. The basic conclusion was that the poorer the student considered his teacher to be, the more economics he learned.

There is little evidence that the ratings/learning relationship involves aspects of learning that are important to the student. Learning implies change, and this can be a trying experience, especially if it involves important attitudes and behavior. It is more comfortable to have one's existing beliefs and skills reinforced. My favorite question for students who are flushed with excitement about a lecture or course is "What is the most important thing that you learned- - something that you can now use and that you did not know previously?" Students take this as a hostile question and seldom provide what they think to be an adequate answer. Typically, they imply that their previous beliefs were reinforced. For example, students give high ratings to outside speakers who tell them to rely on their gut instincts because techniques taught in school are of no value in the real world. I suspect that students rate courses more highly when they confirm their existing beliefs.

Students' ratings of teachers are intended to change the behavior of teachers. We do not have evidence that these changes are likely to contribute to learning. Faculty members with poor ratings might decide that teaching is not rewarding, and spend less time teaching. Teachers might get discouraged by ratings if they see no clear relationship between their attempts to provide a useful learning experience and their ratings. Teachers may get discouraged because time spent on teaching activities has little relationship to ratings, or because, as they develop knowledge in the field through their research, there is no increase in their teacher ratings (Hattie & Marsh 1996).

Faculty members might tailor the class to try to appeal to the least common denominator to avoid having dissatisfied students. At many schools, most teachers are rated “above average” (about 4 on a 5-point scale). Ratings of “1” by disgruntled students can drag a teacher’s average down substantially. Teachers may make their classes less challenging and decide that it is risky to work on skill development. They may give higher grades in the belief that this will improve ratings. They might reduce the work load in the belief that this improves ratings. Some of these beliefs, such as the latter, seem to be correct (Greenwald and Gilmore 1997).

Teacher ratings may reduce experimentation by teachers. Ratings might also lead teachers to abandon approaches that they believe to be most effective for learning. Professors who use experiential learning approaches such as role-playing might find that these are not highly-rated, especially if the exercises provide evidence that students are deficient in important skills.

If one argues that teacher ratings lead to positive changes, then, all other things being equal, teachers who have the most experience and have thus received the most information from the ratings would have the highest ratings. According to Feldman (1983), this is not so.

I believe that teacher ratings are detrimental to students because they are a signal that responsibility for learning lies not with them, but with teachers and administrators. Studies by Tough (1982), Condry (1977) and Armstrong (1983) suggest that when people do not accept responsibility for their learning, they are not very successful. The loss of responsibility is expected to be most serious when the goal is skill development, especially when these skills are important to the learner. The special section provides no evidence on how teacher ratings affect student responsibility.

Why not assess learning directly? Ask the students to answer questions about their own performance because it is the students, not teachers, who are the producers of learning. Students

could be asked “Were you clear about your objectives? Were you well-prepared? Were you organized? Did you spend much time on learning tasks? Did you do the assigned work to the best of your ability? What new concepts and techniques did you master?” Their responses are likely to be related to learning. An even more insightful way to assess learning would be for students to keep learning diaries describing how they applied concepts and techniques to new situations. Besides aiding the learning process, these procedures would provide information to teachers about what the students are learning and how they are learning.

For an even more direct approach, departments (not teachers) could give common examinations, preferably at the end of a program of study. At the start of the program, students could be presented with a list of techniques and concepts to be mastered. If skill training is involved, assessment centers could be used. For example, it should be a relatively straightforward process to design exercises to determine whether students have mastered various techniques, such as how to listen to a client, set objectives, design an experiment, conduct a survey, assess the reliability and validity of information, or use statistical evidence. The effectiveness of Professor X could then be directly assessed by how well her students did in these evaluations. Under the current teacher-ratings culture, many of our graduates have not mastered these basic skills.

Direct third-party measures of learning would reduce my concerns that measurement procedures corrupt the educational process. They might lead faculty members to use methods to enhance learning, such as experiential exercises and learning diaries.

Would it be detrimental for a school to abandon an existing teacher evaluation system? Here again, the special section provides no evidence. However, this seems unlikely for prestigious schools. They derive their reputations from research, not teaching, according to findings in Armstrong and Sperry (1994).

In summary, I expect that teacher ratings will reduce teachers' interest in helping people learn, while reducing student responsibility. These changes will produce detrimental effects on teachers and students such that, over time, learning will decrease and the intellectual environment of a university will be harmed.

Teacher ratings have been more prevalent in recent decades. If they are useful, experienced professors should see improvements in students' learning after introduction of the ratings, or after more emphasis is placed on them. If on the other hand, students are taking less responsibility for their education, professors might observe that students are learning less in an environment that is becoming less learning oriented. This is not systematic evidence, but I have met many professors in the second category and none in the former.

The current political climate will not allow publicly-funded schools to eliminate teacher assessments. However, it should be possible to revise the process so that learning is directly assessed. This would aid students' learning and improve the assessment of teachers. It might also improve the satisfaction of students and teachers.

References

- Armstrong, J. S. (1983), Learner responsibility in management education, or ventures into forbidden research. *Interfaces*, 13 (2), 26-38.
- Armstrong, J. S. & Sperry, T. (1994), Business school prestige: Research versus teaching. *Interfaces*, 24, (2), 13-43.
- Attiyeh, R. & Lumsden, K. G. (1972), Some modern myths in teaching economics: The U.K. experience. *American Economic Review*, 62, 429-433.

- Condry, J. (1977), Enemies of exploration: Self-initiated versus other-initiated learning. *Journal of Personality and Social Psychology*, 35, 459-477.
- Feldman, K. A. (1983), Seniority and experience of college teachers as related to evaluations they receive from their students. *Research in Higher Education*, 18, 3-124.
- Greenwald, A. G. (1997), Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52, 1182-1186.
- Greenwald, A. G. & Gillmore, G. M. (1997), No pain, no gain? The importance of measuring course workload in student ratings of instruction, *Journal of Educational Psychology*, 89, 743-751.
- Hattie, J. & Marsh, H. W. (1996), The relationship between research and teaching: A meta-analysis. *Review of Educational Research*, 66, 507-542.
- Tough, A. (1982), *Intentional Changes*. Chicago: Follett.