

## Key Dimensions of Effective Teaching: What We Can Learn from Studies of Student Evaluations

**William E. Addison**

*Eastern Illinois University*

Effective teaching is obviously a function of multiple factors, a fact that is supported by both empirical research and anecdotal evidence. Like anyone who has taught for a significant period of time, I have had a number of students over the years share their unsolicited views about their instructors that support the multidimensionality of effective teaching. For example, a student might say something like, “My professor really knows the material, but her presentation seems very disorganized.” Or, “He’s a really nice person who seems to care about students, but he just doesn’t do a good job explaining the material.”

Similarly, findings from numerous studies of student evaluations of teaching (SETs) are consistent with the notion that students tend to distinguish among different aspects of effective teaching. For example, in a study I conducted with my colleague John Best about the influence of the “perceived warmth” of the instructor on student ratings (Best & Addison, 2000), we found that higher levels of perceived warmth were associated with higher ratings on items related to class climate (e.g., “Encourages the class to ask questions and share viewpoints”), but were not associated with “skills-related” items (e.g., “Presents the material clearly”). These and many other findings suggest that students do in fact make distinctions among different dimensions of teaching effectiveness. Thus, it is not surprising that well-designed instruments used to assess students’ opinions of teacher effectiveness are multidimensional.

Herbert Marsh, the major figure in the study of student ratings over the last 30 years, has suggested that it is important to consider the purposes of student evaluations when examining the multiple dimensions of SETs. According to Marsh (1991), SETs can provide (a) formative feedback to faculty about their teaching, (b) a summative measure of teaching effectiveness for personnel decisions, (c) information for students to use in selecting courses/instructors, and (d) outcomes that can be used for research on teaching. Most of us are primarily interested in the first two purposes, although as the fourth purpose suggests, SETs have provided a fertile ground for research. In fact, dating back to the 1920s, literally hundreds of studies have been conducted on SETs.

Among the many studies on SETs, a number have employed factor analysis in attempts to distill the various elements of effective teaching into a small number of dimensions. For the purpose of this chapter, I selected four of these studies on the basis of admittedly subjective criteria of historical and substantive importance. I will also admit that I chose these particular studies because the findings tend to be consistent with my own views on the subject.

One of the earliest studies involving factor analysis of student evaluations was published in 1943 by Hermann Remmers and his colleague N. T. Smalzried

(Smalzried & Remmers, 1943). Remmers was a professor of education and psychology at Purdue University beginning in 1923, and he was one of the pioneers in the use and study of student ratings of teaching effectiveness. He designed the Purdue Rating Scale for Instructors, a version of which is still in use at my institution. The current version of the scale is actually a “cafeteria list” that includes 190 items from which instructors can select a subset to form an appropriate evaluation instrument.

In their study, Smalzried and Remmers used an early version of the Purdue scale with just 10 items, each of which was scored on a graphic rating scale. The 10 “traits” that comprised the scale included such obvious qualities as presentation of the subject matter and fairness in grading. They also included some less obvious ones, such as personal appearance and, one of my favorites, “personal peculiarities.”

Although they conducted their study on ratings of high school teachers, their analysis yielded some interesting results that I believe are applicable to all psychology teachers. The analysis resulted in the identification of two factors, which Smalzried and Remmers called the “Empathy Trait” and the “Professional Maturity Trait.” According to the authors, the Empathy Trait can be viewed as “the ability and willingness to wear each student’s sensorial and emotional shoes” (p. 366), and includes such qualities as fairness in grading and sympathetic attitude toward students. The Professional Maturity Trait includes items related to “the tools of the trade” (e.g., presentation of the subject matter).

About 20 years later, Robert Isaacson and his colleagues, including Wilbert McKeachie, conducted a study at the University of Michigan on the dimensions of student evaluations (Isaacson et al., 1964). They collected their data from about 300 introductory psychology students using an instrument that included 145 items, in contrast to the 10-item scale used by Smalzried and Remmers (1943). Their analysis yielded 6 factors: (a) Overload (e.g., assigning a large amount of work); (b) Skill (e.g., explaining material clearly, stimulating intellectual curiosity); (c) Structure (e.g., following the syllabus, planning daily activities); (d) Feedback (e.g., providing comments, pro and con, on students’ work); (e) Group Interaction (e.g., encouraging student participation); and (f) Rapport (e.g., listening attentively to students, providing reasons for criticism). With the possible exception of Overload, most of these factors seem strikingly consistent with one or the other of the two factors identified by Smalzried and Remmers. Specifically, Skill, Structure, and Feedback could probably be subsumed under Professional Maturity, and Group Interaction and Rapport could reasonably be included under Empathy.

Another factor analysis comes from Peter Frey’s 1978 study conducted at Northwestern University. Frey collected data from more than 26,000 student rating “cards” mailed to undergraduate students. Each card included 7 statements on which students indicated their level of agreement. Examples of these statements are “The student had to work hard in this course,” “Class discussion was welcome in this course,” and “The student was able to get personal help in this course.”

Frey’s analysis revealed two factors that he called “Pedagogical Skill” and “Rapport,” which are virtually identical to the Professional Maturity and Empathy traits identified by Smalzried and Remmers (1943). When Frey combined the results from this study with research he conducted earlier, he found that ratings on the “skill” factor were positively related to student learning, but were unrelated to class size and grade. That is, students who performed better on a common final exam rated their instructor more favorably on the skill dimension. Additionally, he found that ratings of “rapport” were positively related to grades and negatively related to class size. Instructors in classes where the average grade was high were rated more favorably on

the rapport dimension, and, not surprisingly, instructors of larger classes were rated lower on rapport.

Given Herbert Marsh's extensive work on SETs, it is appropriate to include one of his studies in this discussion. Marsh (1991) examined data from more than 2000 evaluations, using the 35-item Student Evaluations of Educational Quality (SEEQ) instrument that he designed. After conducting several different factor analyses, Marsh concluded that SEEQ responses could *not* be distilled into one, two, three, or even four factors. This conclusion is not surprising given that previous factor analyses on larger samples had consistently supported a 9-factor structure for the SEEQ (e.g., see Marsh, 1984; Marsh & Hocevar, 1984). These factors are: (a) Breadth of Coverage (e.g., discussed current developments); (b) Organization/Clarity (e.g., objectives stated and pursued); (c) Learning/Value (e.g., course is challenging, stimulating); (d) Examinations/Grading (e.g., exams were fair); (e) Enthusiasm (e.g., dynamic and energetic); (f) Rapport (e.g., interested in individual students); (g) Group Interaction (e.g., encouraged class discussion); (h) Assignments/Readings (e.g., readings were valuable); and (i) Workload/Difficulty (e.g., course workload was light/heavy).

Although clearly less sophisticated than the factor analyses that Marsh used, an informal, "eyeball" analysis of the 9 factors suggests that most of them could be included under the "skill" and "rapport" dimensions seen in the earlier studies. Specifically, Breadth of Coverage, Organization/Clarity, Learning/Value, and Examinations/Grading appear to be skills-related; and Enthusiasm, Rapport, and Group Interaction would seem to be associated with rapport/empathy. I would suggest that the remaining 2 factors, Assignments/Readings and Workload/Difficulty, are more a function of the *course* than the instructor. This admittedly subjective reframing of Marsh's factors supports, to some degree, the notion that there may be two key factors involved in effective teaching: skill and rapport.

Any implications of these studies must be qualified by the limitations of factor analysis, the most salient of which is that the results of these analyses are dependent on the nature of the instrument used, the number of responses, etc. For example, the scale used by Frey (1978) at Northwestern had 7 items, whereas the scale that Isaacson et al. (1964) used at Michigan had 145 items; obviously analyses on these instruments are likely to yield different factor structures.

With the limitations of factor analyses in mind, the overall results from these studies suggest that teachers who are skillful in their presentation of material and who establish and maintain rapport with their students are likely to be viewed by students as effective teachers. Additionally, in light of evidence that students tend to emphasize the rapport dimension over the skill dimension (see Feldman, 1976), teachers who are looking to improve their student ratings should attend to the relationship-building element of teaching in addition to such "nuts and bolts" aspects as knowledge of material, organization of the course, and presentation of the material. As a number of recognized authorities on teaching have reminded us, teaching that includes such qualities as respect for students, availability and helpfulness, and openness to questions is not only good practice because it is the right thing to do, but also because students will respond positively to such behaviors, likely resulting in an increased motivation for learning (e.g., Buskist, Sikorski, Buckley, & Saville, 2002; Lowman, 1995; McKeachie & Svinicki, 2005).

From the evaluator's view, an important implication of these findings is that any interpretation of student ratings based on a single score is likely to be misleading, whether the score is on a "global" item or an average across all items. Similarly, the reliance on single scores will probably yield inconsistent assessments of teaching, a notion that probably helps explain the discrepancies in results from various studies of

SETs. Additionally, the assessment of teaching effectiveness based on a single score may underestimate teaching skill in large classes, and in classes taught by “hard” graders. The reliance on single scores from SETs is obviously a significant issue because such scores are frequently used by administrators and others in making decisions on promotion and tenure.

Ultimately, meaningful assessment of SETs should be done by evaluating the ratings in terms of at least two dimensions. Specifically, ratings could be examined by considering a mean for “skills” items as well as a mean for “rapport” items. This approach would not only provide more useful information for instructors interested in improving their teaching, but it would also yield a more accurate assessment of the instructor’s effectiveness.

## References

- Best, J. B., & Addison, W. E. (2000). A preliminary study of perceived warmth of professor and student evaluations. *Teaching of Psychology, 27*, 60-62.
- Buskist, W., Sikorski, J., Buckley, T., & Saville, B. K. (2002). Elements of master teaching. In S. F. Davis & W. Buskist (Eds.), *The teaching of psychology: Essays in honor of Wilbert McKeachie and Charles L. Brewer* (pp. 27-39). Mahwah, NJ: Lawrence Erlbaum.
- Feldman, K. A. (1976). The superior college teacher from the students’ view. *Research in Higher Education, 5*, 243-288.
- Frey, P. W. (1978). A two-dimensional analysis of student ratings of instruction. *Research in Higher Education, 9*, 69-91.
- Isaacson, R. L., McKeachie, W. J., Milholland, J. E., Lin, Y. G., Hofeller, M., Baerwaldt, J. W., & Zinn, K. L. (1964). Dimensions of student evaluations of teaching. *Journal of Educational Psychology, 55*, 344-351.
- Lowman, J. (1995). *Mastering the techniques of teaching* (2<sup>nd</sup> ed.). San Francisco: Jossey-Bass.
- Marsh, H. W. (1984). Students’ evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*, 707-754.
- Marsh, H. W. (1991). Multidimensional students’ evaluations of teaching effectiveness: A test of alternative higher-order structures. *Journal of Educational Psychology, 83*, 285-296.
- Marsh, H. W., & Hocevar, D. (1984). The factorial invariance of student evaluations of college teaching. *American Educational Research Journal, 21*, 341-366.
- McKeachie, W. J., & Svinicki, M. (2005). *McKeachie’s teaching tips: Strategies, research, and theory for college and university teachers* (12<sup>th</sup> ed.). Boston: Houghton Mifflin.
- Smalzried, N. T., & Remmers, H. H. (1943). A factor analysis of the Purdue Rating Scale for Instructors. *Journal of Educational Psychology, 34*, 363-367.