

## STUDENT RATINGS OF COLLEGE TEACHING: WHAT RESEARCH HAS TO SAY

Lucy C. Jacobs

The use of student ratings of faculty has increased steadily over the past 25 years. Large research universities report 100 percent institutional participation in the collection of student ratings (Hazlett, 1990). Indiana University is no exception. The Bureau of Evaluative Studies and Testing (BEST) at IU processed 133,000 Multi-Op student ratings sheets during the last academic year (1993-94). Multi-Op is only one of the faculty evaluation systems used on campus.

The purpose of student ratings is to provide information that can be used by faculty to improve their courses and their teaching and by administrators to make personnel and program decisions. Research tends to support the validity, reliability, and usefulness of student ratings for this purpose.

Faculty often have questions, however, about the factors apart from teaching performance that might influence the ratings that students give. Recent research has explored the influence of a number of variables on ratings. Overall, this research shows that many factors often assumed to be potential biases do not significantly affect the overall ratings of instructors. This brochure will summarize the most recent research on the effect of a variety of factors on ratings.

### COURSE CHARACTERISTICS

#### FACTORS RELATED:

1. **Class size.** Recent research (Williams and Ory, 1992) found an average correlation of  $-.09$  between class size and various rating items. The negative correlation indicates that smaller classes tend to receive higher ratings, but the negligible size of the correlation indicates that class size is not a very important factor affecting the validity of ratings.  
In an earlier study of five thousand classes, Centra and Creech (1976) found that small classes (<15 students) received the highest ratings; those with 16-35 and those with more than 100 ranked next with equivalent ratings; classes with 35 to 100 received the lowest ratings. They attributed the high ratings of courses with more than 100 students to (1) the tendency for departments to assign instructors who are especially skilled at teaching large groups to those courses; (2) special effort by the instructor to make the class presentations clear; (3) the use of multimedia and other large-class techniques; (4) the use of smaller discussion sections with associate instructors to answer questions and provide help to students.
2. **Discipline.** Instructors teaching certain disciplines tend to receive higher student ratings than instructors in other disciplines. Research has shown that the highest ratings are given to courses in the arts and humanities, followed in descending order by biological and social sciences, business and computer science, and mathematics, engineering, and physical sciences (Cashin, 1992). The lower ratings for math and the physical sciences may be due to the fact that students find these courses more difficult and fast-paced. Cashin (1990) argued that students' quantitative skills are less well developed than their verbal skills; hence quantitative-based courses are more difficult for students and more difficult for faculty to teach. The natural sciences have also experienced such a rapid growth of knowledge that instructors may feel pressured to cover increasing amounts of material in each course, and thus students feel rushed and confused.
3. **Reason for Taking Course.** Students tend to give slightly higher ratings to courses in their major fields and/or to courses that are elective rather than required. Feldman (1978) found a small positive relationship between class ratings and the students' average intrinsic interest (prior subject

interest) in the subject area. Thus, required courses may receive lower ratings simply because students are less interested in them. For this reason, it may be a good idea for faculty to include an item that assesses student interest in the course.

4. **Course Level.** Ratings in higher-level courses tend to be higher than in lower-level courses.
5. **Difficulty Level of Class.** Within a discipline, the courses that are more difficult or have greater workloads tend to receive higher ratings from students. Contrary to popular opinion, easy professors do not necessarily receive high student ratings. Some research shows that students see demanding professors as being better (more effective) than easy professors, hence the higher ratings.

Cashin (1992) correlated rating results from over 100,000 classes. The correlation between the item "I worked harder on this course than most courses I have taken" correlated .44 with the overall composite measure, indicating that working harder in a class was positively related to higher ratings.

#### **FACTORS NOT RELATED:**

1. **Time of day class is taught.** The time of day the class meets has no effect on ratings.

#### **STUDENT CHARACTERISTICS**

##### **FACTORS RELATED:**

1. **Expected Grade.** Positive (but low) correlations have been reported between student ratings and expected grade. That is, students expecting high grades in a course tend to give higher ratings than do students expecting low grades. This relationship between expected grades and ratings is usually offered as evidence for the validity of ratings. Students can distinguish among instructors on the basis of how much they have learned. If they have learned more, they expect higher grades and will give higher ratings. But faculty do not receive high student ratings just because they give high grades.
2. **Motivation.** If students are motivated because of a prior interest in the subject matter or because they chose the class as an elective, instructors are more likely to receive higher ratings in those classes. For example, the mean correlation between the item, "I had a strong desire to take this course" with the overall rating was .39 (Cashin, 1988).
3. **Major.** Closely related to the above is the finding that majors in a class tend to rate instructor/course more positively than nonmajors.
4. **Gender.** The research on the effect of students' gender on the ratings they give has not been conclusive. Early research concluded there was little or no relationship between gender and student ratings. Kierstead and others (1988), however, found that both male and female students consistently rated their female instructors lower than male instructors. Both genders indicated that they had different expectations for female instructors; they were expected not only to be highly competent teachers but also to act in accordance with traditional sex role expectations. They concluded that male and female instructors will earn equal ratings for equal professional work only if the women also display stereotypically feminine behavior. Later research found evidence that students tend to give same-gender instructors slightly higher ratings. In a well-designed study that controlled for course, teacher experience, and class size, Lueck and others (1993) found that male students tended to rate male instructors higher, while female students rated female instructors higher than male instructors. But, overall, there was no significant difference in the ratings given by male and female students.

### **FACTORS NOT RELATED:**

The following student characteristics are not related to the ratings they give instructors:

1. Academic ability
2. Age
3. Class Level (freshman or senior)
4. GPA
5. Personality

### **INSTRUCTOR CHARACTERISTICS**

#### **FACTORS RELATED:**

1. **Faculty Rank.** Regular faculty tend to receive higher ratings than teaching assistants (AI's).
2. **Personality.** Certain personality traits of an instructor may be related to students' overall ratings. But this relationship is influenced more by what instructors do in their teaching than by their personality traits. Research shows that students appreciate instructors who are knowledgeable, warm, outgoing, and enthusiastic (Murray, Rushton, and Paunonen, 1990). These same traits are likely to make the person a more effective teacher, so that students are stimulated to greater achievement and learning. We have seen that if students feel they have learned, they will give higher ratings. So the important factor is not how entertaining and funny the instructor is. "Neither the `stand-up comic' with no content expertise nor the `cold-fish expert' with only content expertise receives the highest ratings consistently" (Braskamp and Ory, 1994, p. 180).
3. **Research productivity.** Research productivity measured by the number of publications is positively but only slightly correlated with student ratings ( $r=.12$ ).

#### **FACTORS NOT RELATED:**

1. Age of instructor
2. Years of teaching experience
3. Gender. Analysis of classroom studies indicates no practical difference in the overall ratings of male and female instructors. In 28 studies, the correlation between gender and overall evaluations of the teacher was .02 (Feldman, 1993).

### **ADMINISTRATION OF RATINGS**

#### **FACTORS RELATED:**

1. **Instructor's Presence in Room.** Ratings are generally more positive if the instructor remains in the room.
2. **Time of Administration.** Ratings given during final exams are lower than those given sometime earlier in the term.
3. **Student Anonymity.** Ratings are not considered as reliable if the rater has to identify him/herself. Signed ratings are usually more positive.
4. **Instructions.** Ratings are more positive if the instructor states that the purpose is for his/her promotion or tenure. Any biasing effect of such factors can be eliminated if all instructors follow the standardized directions for administering the ratings. The directions contained in the Multi-Op manual tell instructors that the rating forms should be given to students the last week or two of classes rather than on final exam day, that students should remain anonymous, that professors

should leave the room, and that no statements be made about the intended use of the ratings.

### ***Open-ended Statements***

In general, the responses students make to the scaled rating statements and the written responses to open-ended questions are similar. The written comments usually provide useful suggestions for faculty improvement as well as insights into the quality of faculty work. In fact, many faculty feel that the feedback they get from the written comments to open-ended questions is more important for their self-improvement than the feedback from rating items. This is especially true if the written comments are collected at midterm instead of at the end of the course. The open-ended questions that are specific and narrow in scope provide the most useful information for faculty improvement.

The disadvantage of the open-ended statements is that students are sometimes reluctant to be candid, because they think they can be identified by their handwriting.

### ***Recommendations***

1. In order to ensure that all student ratings are collected under similar circumstances, faculty should use a standard set of procedures for administering the ratings.
2. Student ratings should not be the only source of data used for personnel decisions. Student ratings should be supplemented by peer observations, alumni ratings, self-evaluations, and portfolios containing descriptions of course materials, teaching methods, innovations, students' pre- and posttest scores, and other evidence of teaching effectiveness.
3. Faculty should learn how to interpret the results of the ratings. Use the university-wide and reference group norms for comparative data and be familiar with the factors that can influence ratings such as size of the class, the particular discipline, and whether it is required or elective for most of the students.
4. Personnel decisions should be based on multiple (at least five) sets of ratings from several different courses and different semesters.
5. Make use of the information provided by the ratings to improve teaching. If weaknesses are evident, the instructor should consult a professional staff member responsible for faculty development. Centra (1994) proposed the NVHM model, which stated that at least four conditions must be fulfilled for student ratings to lead to improvement in instruction: (1) N--instructors must learn something new from them; (2) V--instructors must value the new information; (3) H--they must understand how to make improvements; and (4) M--instructors must be motivated to make the changes and improvements. He also pointed out that the knowledge gained from evaluations is most effective when there is a gap between how students evaluate the instructor and how the instructor evaluates him/herself.

## References

- Braskamp, L.A., and Ory, J.C. (1994). *Assessing Faculty Work*. San Francisco: Jossey-Bass Publishers.
- Cashin, W.E. (1990). Students do rate different academic fields differently. In Theall, M. and Franklin, J. (Eds.), *Student Ratings of Instruction: Issues for Improving Practice*. New Directions for Teaching and Learning, no. 43. San Francisco: Jossey-Bass.
- Cashin and Sixbury (1992) Use cashing (1992)
- Centra, J.A., and Creech, F.R. (1976). *The Relationship Between Students, Teachers, and Course Characteristics and Student Ratings of Teacher Effectiveness*. Princeton, N.J.: Educational Testing Service.
- Centra, J.A. (1994). *Current Issues in Evaluating and Improving College Teaching*. Paper presented at the annual meeting of the American Educational Research Association (AERA) meeting in Atlanta, April.
- Feldman, K.A. (1978). Course characteristics and college students' ratings of their own teachers: what we know and what we don't. *Research in Higher Education*, 9, 199-242.
- Feldman, K.A. (1993). College students' views of male and female college teachers: Part II--evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, 34: 151-211.
- Hazlett, D. (1990). *Evaluation of Teaching Effectiveness in the Association of American Universities (AAU): A Survey of Current Policies and Practices*. Unpublished manuscript, University of Pittsburgh.
- Marsh, H.W., and Cooper, T. (1981). Prior subject interest, students' evaluation, and instructional effectiveness. *Multivariate Behavioral Research*, 16: 82-104.
- Ory, J.C., and Braskamp, L.A. (1981). Faculty perceptions of the quality and usefulness of three types of evaluative information. *Research in Higher Education*, 15, 271-282.
- Williams, R., and Ory, J.C. (1992). *A Further Look at Class Size, Discipline Differences and Student Ratings*. Unpublished manuscript, Office of Instructional Resources, University of Illinois at Urbana-Champaign.