

The posting below looks at some of the abuses and misuses of student ratings of faculty performance. It is from Chapter 4, Uses and Abuses of Student Ratings, by William Pallett in the book, Evaluating Faculty Performance, A Practical Guide to Assessing Teaching, Research, and Service, by Peter Seldin and Associates, Pace University. Anker Publishing Company, Inc. 563 Main Street, P.O. Box 249, Bolton, MA 01740-0249 USA [www.ankerpub.com] Copyright © 2006 by Anker Publishing Company, Inc. All rights reserved. ISBN 1-933371-04-8

## **TOMORROW'S TEACHING AND LEARNING**

### **ABUSES AND MISUSES OF STUDENT RATINGS**

Until quite recently student ratings have been overemphasized and underutilized. When I joined the IDEA Center in 1997, it was common to learn that a campus relied entirely, or almost entirely, on student ratings to assess teaching effectiveness. When asked if student ratings were used to support teaching improvement efforts, the answer was frequently no. As a consequence, faculty at many institutions saw little or no benefit from student ratings. There was some justifiable fear that one composite number might have an adverse impact on one's professional future, so the stakes were high. New faculty members had made substantial investments of both time and money to get where they were. But most of this preparation was not directed at a major component of their responsibilities-teaching.

While many graduate programs have recently placed greater emphasis on teaching, most of a faculty member's preparation to enter the academy is still focused on acquiring disciplinary knowledge. While such preparation is essential, there is a rapidly expanding body of knowledge about how people learn and the uses of technology to support learning that have typically received little attention in graduate school. Therefore, much of what is important to teaching and learning must be learned on the job. It should not be surprising that initial feedback from student ratings is often discouraging.

By the end of their graduate school experience, new faculty have usually received substantial feedback from credible and credentialed graduate professors. When students, whose credibility is suspect, provide negative feedback to them, it is understandable that the results create defensiveness and skepticism. When initial experiences with student ratings are negative and there is no confirmatory evidence from trusted sources, the value of student ratings will often be challenged. The credibility of any process requires trust. One of the best ways to establish trust is to gather and use information appropriately.

Student ratings are neither inherently good nor bad. How they are used determines their value (see Chapter 3). When they are used well, they can be helpful in supporting the agendas for which they are intended. When abused, trust is lost, impact is negative, and something potentially valuable becomes damaging. Examples of how the value of student ratings may be diminished or lost follow.

## Abuse 1: Over reliance on Student Ratings in the Evaluation of Teaching

The IDEA Center has long recommended that student ratings comprise no more than 30% to 50% of the evaluation of teaching (Hoyt & Pallett, 1999). There are a number of components of effective teaching that students are simply not well equipped to judge, including:

- \* The appropriateness of an instructor's objectives
- \* The instructor's knowledge of the subject matter
- \* The degree to which instructional processes or materials are current, balanced, and relevant to objectives
- \* The quality and appropriateness of assessment methods
- \* The appropriateness of grading standards
- \* The instructor's support for department teaching efforts such as curriculum development and mentoring new faculty
- \* The instructor's contribution to a department climate that values teaching

Faculty peers (either local or at a distance) and department/division chairs are much better equipped to address such issues.

No method used to assess teaching effectiveness is perfectly valid, including student ratings.

Because personnel decisions dramatically impact both an individual's personal and professional future and the quality of the educational experience an institution provides, it is vital to use multiple sources of information in assessing all components of effective teaching.

## Abuse 2: Making Too Much of Too Little

While there is substantial evidence that student ratings are reliable, there is always some "noise" in survey data (see Chapter 10). Therefore, if the same student rating survey was administered two days in a row, results would not be precisely the same. Too often, student ratings averages are treated in the same way as things like height and weight that have much less variability over short time intervals. This problem is exacerbated when there are small numbers of raters making judgments, as is the case in classes with fewer than 10 students.

Campus officials often arrive at judgments that make too much of too little. Is there really a difference between student ratings averages of 4.0 and 4.1? Differences in salary increase and other personnel recommendations have often been based on very small differences such as these. To avoid the error of cutting a log with a razor, student ratings results should be categorized into three to five groups for example, "Outstanding," "Exceeds Expectations," "Meets Expectations," "Needs Improvement but Making Progress," and "Fails to Meet Expectations." Utilizing more than three to five groups will almost certainly exceed the measurement sophistication of the instrument being used.

### Abuse 3: Not Enough Information to Make an Accurate Judgment

The IDEA Center recommends ratings of six to eight classes representing all of one's teaching responsibilities be used in the evaluation process-more (eight to twelve) if class sizes are small.

At times people infer from this statement that we recommend rating every class every term, which is not the case. Survey fatigue, a consequence of administering too many surveys in a term, can be an abuse unless those completing the forms are fully committed to the process. A better plan is to rate every class once every three years. For example, classes rated in year one should be rated again in year four.

Given the kind of impact personnel decisions have, both on individual faculty and the institution, it is imperative to collect enough information to inform good judgments. For important decisions such as tenure, promotion, and reappointment, using ratings from only a few classes is not appropriate.

### Abuse 4: Questionable Administrative Procedures

If student ratings are taken seriously by faculty and administrators, it is likely that students will take them seriously as well. In a meeting with students during a recent campus visit, I asked students how conscientious they were in completing the rating form. Their response was-"It depends." They said if the instructor took the process seriously, they did as well. They cited an example where their ratings were made carefully and thoughtfully. A student cited a department that was especially careful and conscientious in the administration of student ratings; faculty told students how important their feedback was to improving teaching and the curriculum and described how past feedback had been used to make improvements. In contrast, the students reported that a tenured faculty member in another department said he had no interest in what they said. In fact, he told them he rarely looked at the results when they were returned to him. Student reaction was predictable-"Why should I care if he doesn't?"

During a campus visit I heard faculty tell how a colleague administered the surveys at a pizza party. On rare occasions, forms have been returned with grease and smudges that led us to question the conditions under which they were collected. In one case I was told that faculty suspected a colleague of removing all negative evaluations before taking them to the department office.

Administrative processes must be created and employed that do not permit tainting the results. Smaller errors and omissions in processes, such as failure to encourage honest and thoughtful responses, also result in a loss of confidence in the information collected. Unless sound administrative procedures are followed, dependable information will not be provided.

### Abuse 5: Using the Instrument (or the Data Collected) Inappropriately

Occasionally, institutions fail to distinguish, or distinguish inappropriately, among the items on a rating scale. On more than one occasion, individuals made comments similar to the following-"While we have 20 items on our ratings form and allegedly all of them are important in the evaluation process only #7 really matters for making personnel decisions." In other cases, the

average of all items may be used to make a judgment about performance without regard to their importance or relevance. An extreme example of this abuse occurred at a campus that found their computer program had included in their summary measure of teaching effectiveness an item about the quality of the rating form. Less extreme abuses occur somewhat frequently as when a global item such as-"Overall I rate this course as excellent"-is given the same importance in a summary measure as less important methods items like-"The instructor encouraged student-faculty interaction outside of class."

#### Abuse 6: Insufficient Attention to Selecting/Developing an Instrument

The tendency for campuses to rely entirely on student ratings to assess teaching effectiveness is rapidly declining. Campuses that take the evaluation of teaching seriously include student ratings as part of larger body of evidence (see Chapter 8). The content of the student ratings tool should be determined by both the functions of the rating program and the content of other sources of evaluative information.

How effective teaching is defined is important in identifying the sources of evidence to use and, if student ratings are included, the content of the instrument. While descriptions of effective teaching in a number of books and articles have consistent themes (Arreola, 2000; Bernstein, 1996; Fink, 2003; Hoyt & Pallett, 1999). Important differences are also present. Without a thoughtful discussion of what teaching effectiveness means on your campus (or department), it is unlikely a student ratings tool will be selected or created that will serve your purposes well.

Decisions about the purposes of the instrument will impact the content (Cashin, 1996) and the length of the instrument. If you want to use student ratings to serve purposes beyond personnel evaluation-to guide improvement efforts, offer descriptive information that assists in advising, or serve as a supplemental source of evidence for accreditation-the instrument will need to be longer than one whose only intent is personnel evaluation.

#### Abuse 7: Failure to Conduct Research to Support the Validity and Reliability of a Student Ratings Tool

When individuals call to inquire about the IDEA student ratings instrument, I usually ask about the student ratings instrument they currently use. Invariably, if it is locally developed, they report having no evidence to support the instrument's validity or reliability. While there are often good reasons to have a locally developed instrument, it is extremely important to establish its credibility through reliability and validity studies. Without such studies, many faculty members (especially those who are psychometrically sophisticated) will lack trust in the instrument. In addition, if a personnel decision is ever challenged in a grievance hearing or lawsuit, those who use the instrument will be on firmer ground if evidence supports the reliability and validity of the system.