Recommendations on the Use of Student Ratings Data for Evaluation of Courses and Faculty

Assessment Resource Team, Antioch University

February 2014

As Antioch University continues its evolution to an integrated university with distinct regional identities, an important line of work is the development of more consistent and utilitarian ways of collecting and using student ratings data. We contend in this white paper that:

> **Student ratings of instruction play a key role in the assessment of teaching effectiveness, courses, and curricula, as long as the design, collection, and use of these data adhere to certain measurement principles, and to the needs and interests of individual campuses and their faculty.**

The paper is organized as follows: 1) we present a brief review of the research on student ratings, their validity, reliability, and usefulness, 2) we draw upon this review to propose several principles that should underlie any student ratings system at Antioch, and 3) we suggest what the core elements of the system should be.

I. Student Ratings: The Research Evidence

The trustworthiness of student ratings of instruction is one of the most-researched topics in all of higher education, and it would take a paper of many pages just to synthesize and summarize it all. Fortunately, this has been done, not just once but several times. Two of the most comprehensive reviews of the student ratings literature are:

- "Evaluating Teacher Effectiveness: Research Summary," by Terry Doyle, Center for Teaching, Learning, and Faculty Development at Ferris State University (http://learnercenteredteaching.wordpress.com/articles-and-books/evaluating-teacher-effectiveness-%E2%80%94research-summary/) (2004), and

- "Student Ratings of Teaching: A Summary of Research and Literature," by Stephen Benton and William Cashin (2012), updating an earlier meta-analysis from 1995.

Together, these two articles address most of the concerns faculty and administrators might have about student ratings of instruction. Key conclusions from these reviews highlight *misconceptions* about student ratings, none supported by research,

such as: students' ratings are more like popularity contests, invalid and unreliable, leading to grade inflation, and that students want easy courses, make inconsistent judgments, and are unappreciative of good teaching.

On the contrary, with certain caveats to be noted later, research indicates that student ratings are generally:

1. Valid as indirect measures of student learning.
2. Reliable (i.e., consistent across raters, courses, and time periods).
3. Free from most sources of extraneous bias, such as class size, grading standards, time of day, and demographic variables such as faculty race, gender, rank, age, years of experience, and personality traits.
4. Minimally impacted by student characteristics, such as gender or academic ability.

Furthermore:

- The instructor, not the course, is the primary determinant of student ratings, although students do rate elective courses higher than required courses, and advanced courses higher than introductory ones.
- Teaching characteristics most related to ratings of overall effectiveness include stimulating student interest, fostering student collaboration, establishing rapport, encouraging student involvement, and structuring the classroom.
- Very slight differences in ratings exist according to discipline, with arts and humanities receiving the highest ratings, followed in order by biological and social sciences, business, computer science, math, engineering, and the physical sciences.
- Contrary to popular belief, student ratings are *positively* correlated with course difficulty and required student workload, as long as the work required is seen as contributing to course objectives.
- Student ratings of the "amount learned" are moderately correlated with overall ratings of the instructor and the course. However, faculty have limited control over the factors that affect the amount students learn.
- Consistently small relationships have been found between student ratings and students' prior interest in the subject matter and whether the course is part of the student's major area of study.
- Ratings requiring student signatures are slightly higher on average than anonymous ratings. Ratings are also higher when students know that the ratings may be used for faculty promotion or are told by the administrator of the instrument how important the feedback will be.
- Ratings by students show consistently moderate correlations with ratings by others, including supervisors, peers, and trained observers.

There has been a greater reliance in recent years on online administration of student ratings surveys. Benton and Cashin (2012) reviewed the available evidence about online vs. paper surveys and found that response rates to open-ended items tend to be higher, and written comments lengthier, in online forms. Response rates to the "fixed" items are however lower, which leads to concerns about how representative those who respond are of the total class. Despite the lower response rates, studies comparing paper vs. online forms have found little if any evidence of response bias (e.g., disproportionate percentages of students with negative views).

A related question is the suitability of standard student ratings instruments for online courses. Findings reported above generally hold true for online courses, although response rates to the student ratings instrument do tend to be lower.

Among those who study student ratings in higher education, there is virtual consensus on one point: despite the overall validity, reliability, and potential usefulness of student ratings, they should *never* be used as the sole source of data for evaluating teaching effectiveness, but should instead be part of a process that includes other sources of evidence, such as peer review and self-evaluation. Students are qualified to evaluate some aspects of instruction, such as perceived fairness, overall satisfaction, and the extent to which the course resulted in valuable learning. They are generally *not* qualified to address such things as faculty knowledge of the subject matter, or whether the content covered was current or appropriate for the course.

Finally, research on student ratings suggests that the effectiveness of any student ratings system rests on much less the content of the form than on: having clear purposes for which the ratings will be used; helping faculty understand the uses and limitations of student ratings; educating students on how to give useful feedback; and helping those who will interpret the data to do so effectively and fairly. It is also the case that instructors benefit most from student ratings when they have an opportunity to influence the questions that are asked and when resources for improving their courses and practices are made available (Gaubatz, 2000).

II. Principles for a Course Evaluation Process at Antioch

Collecting student ratings data must be characterized primarily as a way to answer questions that a faculty member or program want to address. Thus, developing rating instruments for course evaluations should be driven by faculty who have specific assessment questions, either for a particular course or for a program's curriculum as a whole. If the question is "how can we improve the effectiveness of a course?" then the elements of the rating instrument must be crafted to directly query students about the way a particular course was delivered and how it met expectations.

In addition, student ratings data can address questions such as "does this course help students to meet the program learning outcomes?" or "do the cumulative data generated by courses across a program and over time indicate that the program is fulfilling its goals?" Program directors and other academic administrators may benefit from summary reports of student ratings to address questions related to quality assurance and faculty accountability.

Thus, any course evaluation process for Antioch University must be able to address a range of key questions, with the first priority to inform the faculty teaching the courses followed by the program whose goals the courses were designed to serve. In the main, core principles for student ratings at Antioch should be founded on the research consensus about good practice:

- Any student ratings system should be part of a program's overall assessment plan. Decisions about the content and use of a student ratings form should not be separate from discussions about curricular goals, markers of effective teaching in the program, and sources of evidence (including student ratings) used to make judgments of teaching quality.
- The content of a student ratings form should focus on information useful to the program and individual faculty. All stakeholders (students, faculty, administrators) must have a clear sense of how the data will be used.
- Students need to be trained on how to give meaningful feedback and also understand the purpose of the ratings. The "Closing the Loop" element of the assessment process cannot be invisible to students; demonstrating how the data they provide are used to improve practice increases student investment in the process.
- Relatedly, the form should have the flexibility for faculty members to add diagnostic questions of their own choosing, with responses to these questions seen only by them.
- Faculty and administrators using the data must be trained on how to interpret the results.

The above principles lead to this one:
While the University may adopt a central technology to be used by all programs, there should not be a single University (or campus) form, with only one set of items to be used by all. With the exception of certain "core elements" enumerated below, content should be determined first at the program level, by the faculty.

III. Core Elements of the System

   While flexibility and faculty/program ownership is essential, we do believe that a student ratings instrument should contain three kinds of items, irrespective of discipline or profession:

1. **Items linked to congruence of course with program goals and student learning outcomes (thus allowing students to validate the mapping of curriculum to the program's intended goals and outcomes);**
2. **Items linked to student accomplishment of course objectives; and**
3. **Items about instructor fairness and respect for students, and the degree to which the instructor was responsive to student needs.**

   A final core element should be **clear, university-wide policies regarding access to evaluation data.** Assuming that the University adopts a single system (i.e., Blue) for collecting, analyzing, and reporting student ratings data – which appears to be capable of adhering to the principles and core elements listed above – Antioch will have to establish firm policies on balancing standardization with local flexibility and confidentiality.

   As noted earlier, content should be largely determined at the program level in order to answer the questions program faculty have about their courses. Each form should have two sections, one containing standard items agreed-upon by the program, the other diagnostic items chosen by the faculty member. Only the faculty member should have access to both sections; a separate report containing results of the first section should go to the program director and possibly the CAO.

Respectfully submitted,
Joe Cronin, AUM   Catherine Radecki-Bush, AUSB
Suzanne Engelberg, AUS   Andrea Richards, AULA
Tom Julius, AUNE   Jon Wergin, AUL&C

References
Benton, S. L., and Cashin, W. E. (2012). *Student ratings of teaching: A summary of research and literature.* Kansas State University: The IDEA Center, *IDEA Paper #50.*

Doyle, T. (2004). *Evaluating teacher effectiveness: Research summary.* Ferris State University: http://learnercenteredteaching.wordpress.com/articles-and-books/evaluating-teacher-effectiveness-%E2%80%94research-summary/.

Gaubatz, N. (2000). "*What's the Use of Student Ratings of Teaching Effectiveness?*" http://csti.syr.edu/csti/T-L/stdrate.htm 2000.