

Student Ratings versus Student Evaluations of Teaching

[Raoul A. Arreola, Ph.D.](#)

Professor Emeritus, The University of Tennessee Health Science Center

One of the core issues that must be addressed in considering the process of designing a valid and reliable instrument to be completed by students as part of the assessment of faculty teaching performance is the differentiation between student *ratings* and student *evaluations*. Key to understanding the difference is to examine the definition of the term *evaluation*:

Evaluation is the process of interpreting a measurement (or aggregate of measurements) by means of a *specific value construct* to determine the degree to which the measurement(s) represent a desirable condition. The result of an evaluation is a *subjective judgment* as to whether the performance conforms to the value construct and is thus ‘desirable’ or is at variance to the value construct and is thus ‘undesirable’.

Considering this definition of evaluation, and examining the role of student input into the evaluation of teaching performance, we can take one of two directions; 1) construct a form that asks students to make judgments as to the desirability of certain instructor performances or characteristics (i.e. evaluate teaching and the teacher), or, 2) construct a form that measures the students’ perception of the impact of the instructor’s performance on their learning students and from which a judgment as to the desirability of the performance (evaluation) *will be made later by the systematic application of a specific, predetermined value construct*.

Taking the first approach results in constructing *student evaluation* forms. In such forms students are asked such questions as:

“How would you rate this instructor compared to all the other teachers you have had?” or,
“Overall, how would you rate this teacher?”

With such questions the student is asked to classify the faculty member in different ways such as ‘among the top 10%’, ‘exceptional’, ‘unsatisfactory’, ‘among the worst’ and so on. The problem with this approach is that the judgment as to whether the faculty member’s performance represents a desirable or undesirable condition is being made on the basis of the student’s *individual value construct derived from their personal (unknown) experience*. Each student’s individual value construct may not only be unknown to either the faculty member or the administrator reading the results, *but may differ from student to student*.

For example, one student in a class may evaluate the instructor as being “Among the BEST I’ve ever had” while another student in class evaluates the instructor as being “Among the WORST I’ve ever had”. Does this mean the instructor is good or poor? The first student may have had educational experiences that involved many bad teachers and this instructor is seen as simply being not as bad as the others. The other student may have had educational experiences that involved many excellent teachers and this instructor is simply the least ‘excellent’ of the bunch.

It is for this reason that student “evaluation” forms (that is, forms designed *only* to gather student evaluative judgments concerning faculty performance) may be considered inherently flawed.

The reason for this is that no matter how technically correct the items may be in terms of their psychometric characteristics, there is no assurance that the value construct being applied by the student to interpret their observations and experiences (and thus make evaluative judgments) either matches that of the faculty or is even consistent from student to student. It is from this flaw in many of the forms used in higher education that the common conception of student evaluations being ‘just a popularity contest’ originates.

As contrasted to student *evaluation* forms, student *rating* forms take the approach of measuring student *perceptions* and *reactions* rather than ask for an evaluation’ of some aspect of teaching. That is, the forms measure those psychological phenomena that the student is best qualified to provide – their own reactions. Thus, well-designed student rating forms will include items such as:

“The objectives of the course were clear,”

“The audio-visuals used in class helped me to learn the material”, or

“The instructor’s speaking style made it difficult for me to understand what he/she was saying.”

Students are then provided a “Strongly Agree -Strongly Disagree” or similar scale with which to respond to each item. In this way, measures of the students’ perceptions and reactions to certain aspects of the instructor’s performance are gathered. The evaluation of the instructor’s performance is then made by applying an agreed-upon value construct to interpret the students’ reactions as being indicative of either a desirable or undesirable teaching performance.

Unfortunately, the terms ‘student ratings’ and ‘student evaluations’ have come to be used interchangeably in higher education, and even in the professional literature. Not all so-called ‘student evaluation’ forms truly call for the student to make an evaluative judgment, but are, in fact, student rating forms (although they may contain one or two items calling for an evaluative conclusion). On the other hand, not all student rating forms are designed to simply measure student perceptions and reactions, but include many items that call for students to come to an evaluative conclusion. It is likely that this mixture of terms and approaches (i.e., measurement of student perceptions and reactions versus requests for students to draw evaluative conclusions) that has led to much of the confusion and continuing concern over the use of student input in the evaluation of faculty performance.

In any case, as has been described in greater detail elsewhere, the evaluation of teaching performance must be based on multiple sources of data, not just student rating data which, of necessity, can only provide part of the picture of the complex (meta) professional performance we call teaching.