

Interpreting Adjusted Ratings of Outcomes

April 2002
(modified December 2008)



The IDEA system provides a number of measures of teaching effectiveness. Chief among these are measures related to *student learning*. Instructors indicate which of 12 learning objectives are relevant (“important” or “essential”) to the class; and students rate their progress on these objectives. An overall measure (*Progress on Relevant Objectives*) is obtained by combining progress ratings on all “important” and “essential” objectives, giving double weight to those considered “essential.” Thus, there are 13 possible outcome measures related to student learning; but only a few of them (the “important” and “essential” objectives) will be used in a given class.

In addition to *Progress on Relevant Objectives*, there are three other “global outcome” measures:

- 1) Increased Positive Attitude (average student response to the item, *As a result of taking this course, I have more positive feelings toward this field of study*);
- 2) Excellence of Teacher (average response to the item, *Overall, I rate this instructor as an excellent teacher*); and
- 3) Excellence of Course (average response to the item, *Overall, I rate this course as excellent*).

Research has confirmed the common belief that instructional outcomes are influenced by factors not directly under the control of the instructor. For example, outcomes tend to be more favorable when classes are relatively small, and when both student motivation and effort are high. To the degree that these factors both affect outcome measures and are beyond the instructor’s control, they constitute relevant “extraneous circumstances.” When these factors are favorable, student ratings tend to be more favorable; when they are unfavorable, the opposite is true.

Although it is always important to estimate the amount of student learning in a given course, when the focus is on the *evaluation of teaching effectiveness*, it is important to separate the contributions of the teacher from the contributions of extraneous factors to student learning. That is the purpose of the IDEA system’s **adjusted ratings**. Adjusted ratings use statistical techniques (regression equations) to exclude the effect of relevant extraneous factors from outcome measures;

thus they provide a relatively pure estimate of the instructor’s impact.

In deciding which ratings to use, it is important to consider whether the focus is on student outcomes or on instructor contributions to those outcomes. For the former, “Unadjusted” ratings are most relevant; for the latter, “Adjusted” ratings are generally more appropriate.

The number of relevant extraneous factors is not known. At the present, the IDEA system assesses five such measures¹, but our research program includes the search for additional variables in an effort to ensure an “even playing field” for faculty members whose teaching effectiveness is judged, in part, by IDEA results. The Diagnostic Form adjustments, comprised of the following five factors, account for 12-24 percent of the variation in the ratings of the 12 learning objectives.

Diagnostic Form

1) Work Habits (mean of Item 43, *As a rule, I put forth more effort than other students on academic work*) is generally the most potent predictor. Classes that contain students who typically worked hard on their studies are expected to receive favorable ratings. Unless ratings are adjusted, the instructors of such classes would have an unfair advantage over colleagues with less dedicated students.

2) Course Motivation (mean of Item 39, *I really wanted to take this course regardless of who taught it*) is the second most potent predictor. Similar to the Work Habits factor, classes with students who are highly motivated to take the course regardless of who taught it are expected to receive favorable ratings. Again, unless ratings are adjusted, the instructors of such classes would have an unfair advantage over colleagues with less motivated students.

3) Size of Class is determined by instructor reported enrollment. The regression coefficient for this factor was not always statistically significant; but when it was, it was always negative – the larger the class, the lower the expected rating. Those teaching small classes have

¹ The Diagnostic Form adjusts for five factors; the Short Form currently accounts for three.

an advantage over those teaching large classes; hence, in the interest of fairness, ratings should be adjusted to take this into account.

4) Course Difficulty, as indicated by student ratings of item 35, *Difficulty of subject matter*, can not be used as a direct measure of a relevant extraneous factor because the instructor influences perceptions of “difficulty” (for example, by assignments or by expectations of intellectual effort). A statistical technique was used to remove the instructor’s influence on “Difficulty” ratings in order to achieve a measure of a class’s (and often a discipline’s) inherent difficulty. Generally, if the class is perceived as difficult (after taking into account the impact of the instructor on perceived difficulty), an attenuated outcome can be expected. This was especially apparent in progress ratings on “Creative capacities” and “Communication skills” where high difficulty is strongly associated with low progress ratings. The two exceptions, where “disciplinary difficulty” had a positive effect on the predicted outcome, are for the progress ratings concerned with basic cognitive development (“Factual knowledge” and “Principles and theories”). In these instances, conventional wisdom (high difficulty = low ratings) is solidly refuted.

5) Student Effort is measured with responses to item 37, *I worked harder on this course than on most courses I have taken*. However, for most students, responses reflect both the student’s general academic work habits and the teacher’s success in stimulating effort. Since the latter is under the teacher’s control, it was removed (statistically) from ratings. The residual is considered a measure of our fifth extraneous factor: student effort not attributable to the instructor. In most cases, student effort in the class is also negatively related to ratings. Classes containing an unusually large number of students who worked harder than the instructor’s approach required ended up with lower progress ratings. The reason for this is unclear. Perhaps those who found it necessary to put in extra effort were those whose backgrounds did not prepare them well for the class. They may also be students who lack self-confidence and, for this reason, under-achieve (or under-estimate their progress in a self-abasing manner).

Adjustments for the three global ratings merit special scrutiny. Regression results for predicted scores on “Increased positive attitude” and “Excellent course” were similar to each other. The order of the most influential predictors was reversed over that found for individual progress ratings; Course Motivation was the clear leader, and Work Habits was a relatively distant second. Classes perceived as very difficult tended to be rated low on these measures, but (again in contrast to the findings for individual progress ratings) those with students who worked unusually hard tended to be rated

more favorably. Scores will be adjusted downward for classes with students who are highly motivated and dedicated, especially if the class was perceived as unusually difficult.

The measures of extraneous circumstances did not have as much influence on the adjustment for the other global rating (“Excellent instructor”). Although significant regression weights were found for all five independent variables, these were all of modest magnitude. Course Motivation and Work Habits were about equal in their influence on such ratings, while the adjusted ratings for “Difficulty” and “Effort” had a more moderate (and negative) influence. Enrollment size had a very minor and negative influence. Thus, instructor “popularity” was not accurately predicted by these measures; but student motivation and dedication did have a moderate positive influence while disciplinary difficulty and student effort had a slight negative influence.

Short Form

Adjustments to ratings on the short form are less precise because it provided no information on Disciplinary Difficulty or Student Effort. The current Short Form adjustments are comprised of the following three factors:

1) Course Motivation (item 15, *I really wanted to take this course regardless of who taught it*) is the most important extraneous variable taken into account by adjustments to the Short Form; the stronger the desire of enrollees to take the course regardless of who taught it, the more likely high progress ratings would be reported. Student attitudes clearly influence student learning.

2) Size of Class, as defined by instructor reported enrollment, does not always have a significant regression weight. When it does, the weight is negative. If classes are large, it is probable that progress ratings would be negatively affected; in these cases, adjusted scores will be higher than actual scores.

3) Work Habits (mean of Item 13, *As a rule, I put forth more effort than other students on academic work*) is generally the most potent predictor. Classes that contain students who typically worked hard on their studies are expected to receive favorable ratings. Unless ratings are adjusted, the instructors of such classes would have an unfair advantage over colleagues with less dedicated students.