

Using IDEA Results for Administrative Decision-making



A u g u s t 2 0 0 4

Revised February 2005

- ◆ How much emphasis will you give student ratings in the evaluation of teaching?
- ◆ How many classes need to be rated?
- ◆ Performance categories: How many should you use?
- ◆ Do you plan to authenticate objectives?
- ◆ Will you weight the importance of items or groups of items in the decision-making process? If so, how?
- ◆ Will you use averages on a five-point scale or converted averages?
- ◆ Will you use adjusted or unadjusted scores?

These questions merit discussion on any campus where student ratings are considered in administrative decision-making. Some address general topics related to using student ratings. Others are specifically related to the IDEA Student Ratings of Instruction System. In some cases The IDEA Center makes specific recommendations; more often, options available to you are described.

The most important criterion for successful use of student ratings data is the trust that faculty and administrators have in the system. The IDEA Student Ratings system is intended both to assist in arriving at a summary judgment about effectiveness and to serve as a diagnostic tool for guiding improvement. *Only pages one and two of the IDEA report should be employed in the administrative decision-making process.* However, a few campuses have identified other items of special relevance to an institutional purpose rather than for the intended purpose for which the form was designed. This seems appropriate if the rationale for selecting these items is supported by those involved in the evaluation process. But without such support, it is recommended that items be used only within the confines of the original intent of the instrument.

How much emphasis will you give student ratings in the evaluation of teaching? *The IDEA Center strongly recommends that additional sources of evidence be used when teaching is evaluated and that student ratings constitute only 30% to 50% of the overall evaluation of teaching.* This recommendation is made for two reasons: 1) some components of effective teaching are best judged by peers and not students, and 2) it is always useful to triangulate information (determine the degree to which sources intended to address the same question yield consistent results). The weight to be given to student ratings is dependent on the amount and quality of the other sources of evidence. The more confidence you have in them, the less reliance you should place on student ratings; but, if the student rating system is a sound one

of documented reliability and validity, it should count for at least 30% of the overall evaluation of teaching.

How many classes need to be rated? When using IDEA for making important evaluation decisions, *the Center recommends using six to eight classes*, not necessarily all from the same academic year, that are *representative of all of an instructor's teaching responsibilities*. More classes should be used if class sizes are small (less than ten enrolled). This ensures that no one class will have an undue effect upon conclusions. For example, if you include eight classes and student ratings count 40% in the overall evaluation of teaching, each class would count only 5%.

Performance categories: How many should you use? *The IDEA Center recommends classifying teaching effectiveness into one of three to five categories (e.g., "Failed to meet expectations," "Met expectations," "Exceeded expectations," and "Outstanding performance").* Although there is considerable evidence of the validity of the IDEA student ratings system, no system is perfectly valid. The same is true of other sources of information. Therefore, fine discriminations cannot be justified; there is always "measurement noise." Some campuses categorize performance on the basis of each source of information; others create only an aggregate category. In either case, the use of more than five categories is seldom justified.

In the discussion below, you will see examples of performance categories defined by quantitative measures. Some users prefer a qualitative approach. This requires the development of general guidelines that identify characteristics associated with each classification, but quantitative cutoff points are not used. Typically, the process employs a peer review committee charged with reviewing all evidence of teaching effectiveness and arriving at a consensus regarding the appropriate category. Whatever approach is used should be consistent with the

institutional/departmental culture and be trusted by those involved in the process.

Do you plan to authenticate objectives? Since the IDEA system relies heavily on student ratings of their progress on the objectives chosen by the instructor, it is essential that these choices be made thoughtfully. On some campuses, the decision about objectives is entirely the instructor's prerogative. On others, an "authentication" process is employed to ensure that the objectives chosen are appropriate and consistent with the needs of the institution and/or the curriculum. Discussions with the department chair or the curriculum committee may be held either prior to the completion of the FIF, or after the course has been evaluated. In the latter case, the discussion focuses on why certain objectives were selected. Some campuses authenticate objectives only for multiple section courses in the belief that objectives for such a course ought to be the same for all sections.

There are two reasons some campuses opt to employ an authentication process. Some believe that the IDEA system is susceptible to manipulation in the selection of objectives. For example, if the curriculum committee believes the course should address three of the twelve IDEA learning objectives, but the instructor selects only the one where he/she is likely to receive the highest ratings, ratings may well be higher than would be merited if all relevant objectives were selected. The second reason for "authenticating" objectives is related to curriculum review. In designing the curriculum, it is important to know how each course contributes to program goals. Some departments/curriculum committees identify objectives for each course in the department; others provide instructors the opportunity to add one or two additional objectives.

The IDEA Center makes no recommendation about whether or not to authenticate objectives or what process to use. These decisions should be made to be consistent with the campus and departmental climate.

Will you weight the importance of items or groups of items in the decision-making process? If so, how? Page one of the IDEA report includes the results for two overall evaluation items ("Excellent Instructor" and "Excellent Course") and the combined score of progress ratings on objectives selected as "Important" or "Essential." We call this combined score "PRO" (Progress on Relevant Objectives). Most campuses use these three measures to assess teaching effectiveness. A few users review individual results for each objective identified as "Essential" or "Important," found on page two, rather than the PRO measure that combines them; but most campuses use only the items found on page one for administrative decision-making.

It is important to consider how much each of these three items should contribute to the overall evaluation. Page one of the report contains a "Summary Evaluation" that weights PRO 50%, Excellent Instructor 25%, and Excellent Course 25%. The IDEA Center believes that this way of weighting the three measures is appropriate; it is also in common use among institutions participating in the IDEA program. However, several campuses employ a different weighting scheme or process for arriving at a summary judgment. It is important to have a thoughtful discussion about how best to combine these three measures to best reflect the philosophy and priorities of your institution.

Will you use averages on a five-point scale or converted averages? All of the items on the IDEA survey are rated on a five-point scale. The IDEA report contains both *5-point averages* based on this scale and *converted averages* (standardized scores with an average of 50 and a standard deviation of 10.)

5-Point Averages. Campuses that use a "criterion-referenced" approach to evaluation must use raw averages. An index of teaching effectiveness is called "criterion referenced" if its interpretation is based on pre-established judgments of the meaning of a given average. Any or all of the three summary measures shown on Page 1 (Progress on Relevant Objectives, Overall Ratings, and Summary Evaluation) become "criterion referenced" if the institution establishes standards for describing degrees of excellence that don't rely upon a comparison with results for other instructors or classes. An example is given on the next page.

A true criterion-referenced system creates the cutoff points for various categories (Excellent; Unsatisfactory; etc.) by reviewing the words associated with a given rating (e.g., 1=No Apparent Progress; 5=Exceptional Progress in the IDEA system). In reality, even when categories are criterion based, normative (comparative) information usually influences how categories are determined. Normative results (national, disciplinary, institutional) may be reviewed before categories are determined. An evaluation system will not serve constructive purposes if an inordinately high percentage of faculty members are in either of the extreme categories. If large numbers are in the lowest two categories, energy will be wasted fighting the evaluation system rather than addressing more important questions concerning teaching and learning. On the other hand, if you fail to consider norms in establishing categories, and students follow national trends toward leniency, most faculty members might end up being classified as "Outstanding," diminishing the status of your superior faculty.

Some believe 5-point averages should be used because they prefer a system that de-emphasizes an unhealthy competitive atmosphere; others support the use of 5-point averages on the grounds that the quality of teaching should be judged by the amount of progress students report on objectives stressed by the instructor; they believe that, even though these ratings are generally higher for some objectives than others, this simply indicates that teaching is generally more effective in classes where such objectives are chosen. Individuals who hold this view do not advocate the use of “converted” scores.

Converted Averages. Institutions that want to make judgments about teaching effectiveness on a comparative basis should use **converted scores**. The results for a given class are compared with those for three groups: 1) all classes in the IDEA database (always reported), 2) classes in the same discipline and 3) all classes at your institution. Disciplinary and institutional comparisons are provided if at least 400 classes were rated.

One reason for using converted scores is that the 5-point averages of progress ratings on “Essential” or “Important” objectives are not the same for each. In classes where *Gaining factual knowledge (terminology, classifications, methods, trends)* was an important or essential objective, the average student rating of progress was 4.00; the comparable rating for classes choosing *Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature, etc.)* as an objective was only 3.69. If only 5-point averages were considered, those choosing the “broad liberal education” objective would likely be disadvantaged. The use of converted scores removes the possibility of advantaging (or disadvantaging) oneself solely on the basis of the learning objectives selected. In the example above, the 5-point average score for *Factual Knowledge* of 4.00 is converted to a score of 50; and the average score of 3.69 for the *intellectual/cultural activity* is also converted to 50. This approach ensures that instructors choosing objectives where average progress ratings are relatively low will not be penalized for choosing objectives that are particularly challenging or that address complex cognitive skills.

Converted scores also provide normative information. For each of the three norm groups (national, discipline, institution), the average rating is always 50 and the standard deviation is always 10. Thus, a converted score of 60 is always one standard deviation above the average (i.e., roughly, in the upper 15% of classes included in the norm group). The table provides examples of criterion and normative applications.¹

Criterion Average Rating	Effectiveness Category	Normative T Score Category
Below 3.0	Below acceptable standards	Below 38
3.0-3.4	Marginal, improvement needed	38-44
3.5-3.9	Good	45-54
4.0-4.4	Excellent	55-62
4.5 or higher	Outstanding	63 or higher

Will you use adjusted or unadjusted scores? Some factors over which the instructor has no control influence student ratings. Adjusted averages are used to take some of these “extraneous” influences into account. Adjusted ratings are intended to “level the playing field” across different classes by recognizing that conditions beyond the instructor’s control can increase or decrease student ratings. The IDEA Diagnostic Form adjusts for student motivation, student work habits, class size, course difficulty, and student effort. The Short Form adjusts using the first three factors—motivation, work habits and size of class. If one has a class with poorly motivated students (low average item 39, diagnostic form) with poor work habits (low average on item 43), the adjustment will be upward. If one teaches a class with highly motivated students who have good work habits (high averages on both items), evaluation measures will be adjusted downward. The intent is to “level the playing field,” especially for faculty teaching classes enrolling poorly motivated students with poor work habits. The evidence is compelling that student characteristics have substantial impact on students’ rating of their learning (see IDEA Technical Report 12, page 40). *Therefore, in most circumstances, The IDEA Center recommends using adjusted scores. However, we recommend using the unadjusted score if the average progress rating is high (for example, 4.2 or higher¹).* In these instances, a downward adjustment occurs not because the instructor was less effective than suggested by unadjusted ratings but because extraneous factors (high motivation, good work habits) played such a large role in fostering student achievement that the teacher’s opportunity to influence progress was reduced. In these cases, instructors should not be penalized for having success with a class of highly motivated students with good work habits.

¹Determining specific cutoffs and categories require careful campus discussion. The examples are not intended to be adopted for use.