Interpretative Guide: IDEA Short Form Report

N o v e m b e r 2 0 0 4

This three-page Interpretative Guide provides information to help you understand your IDEA Short Form Report. You will find links in each section that provide expanded details about various topics related to IDEA and the use of student ratings.

Effective teaching is a complex art. It requires sensitivity to the unique objectives of the course, the personality and preferred communication/interaction style of the instructor, the background and motivation of the students, and the peculiarities of the discipline. It is these factors and their interactions that determine the degree to which desired outcomes are achieved. Although student ratings cannot provide **all** of the information needed to evaluate and improve instruction, this guide will help you make more complete and accurate interpretations of results from the *IDEA Short Form Report*.

The *IDEA Short Form Report* is designed to respond to four questions:

1. Overall, how effectively was this class taught?
2. How does this compare with the ratings of other teachers?
3. Were you more successful in facilitating progress on some class objectives than on others?
4. Do some salient characteristics of this class and its students have implications for instruction?

Two kinds of scores are reported: "Average" scores are based on a 5-point rating scale, while "Converted" scores all have an average of 50 and a standard deviation (measure of variability) of 10. Both "Average" and "Converted" scores are presented in "raw" (unadjusted) and "adjusted" forms. Each type of score is important to a complete understanding of your results.

➢ More on Types of Scores

Question 1. Overall, how effectively was this class taught? (Refer to the tables and graph reported on Page 1 of the *IDEA Short Form Report*.)

One of the best ways to infer teaching effectiveness is to examine student ratings of progress on objectives chosen as *Important* or *Essential* by the instructor. The **average** of these rating provides a good indication of how successfully objectives were reached, especially if at least 10 students provided ratings and if at least 75% of enrollees responded.

Progress ratings are made on a 5-point scale: 1=No apparent progress; 2=Slight progress; 3=Moderate progress; 4=Substantial progress; and 5=Exceptional progress. In interpreting "raw" and "adjusted" averages, these terms can be substituted for the numeric figures; e.g., an average of 4.0 indicates that "substantial progress" is an appropriate term for summarizing student ratings.

An overall index of teaching effectiveness (PRO=Progress on Relevant Objectives) combines ratings of progress on the objectives identified by the instructor as *Important* (weighted "1") or *Essential* (weighted "2")[1]. The IDEA Center regards this as its single *best estimate of teaching effectiveness*. Raw and adjusted PRO scores are provided for converted averages as well as for those based on the 5-point rating scale. Converted averages are preferred when making comparisons among faculty members or classes because they take into account the fact that average progress ratings are much higher for some objectives than for others; that is, some objectives appear to be more easily achieved than others. Converted scores assure faculty members that they will not be penalized for selecting objectives that are especially difficult.

Two additional overall measures of teaching effectiveness are shown on the report. These are the average ratings of two items using a 5-point scale (1=Definitely false; 5=Definitely true):

[1] Ratings of progress on individual objectives are provided on Page 2 of the report and can address Question 3.

1. *Overall, I rate this instructor an excellent teacher.*
2. *Overall, I rate this course as excellent.*

As an index of teaching effectiveness, the average of these two ratings is commonly regarded as about equal in value to the "Progress on Relevant Objectives" index described above. Therefore, the **Summary Evaluation** reported on Page 1 averages the PRO score with the average of these two ratings. Although many IDEA users find this method of arriving at a Summary Evaluation to be meaningful, some may feel that other methods for arriving at a summary judgment better reflects their institution's philosophy and/or priorities; they are encouraged to define a process or use an index that best reflects the local situation.

Question 2. How do your ratings compare with those of other teachers? (Refer to the comparisons shown on the right hand side of Page 1 of the *IDEA Short Form Report.*)

Criterion-referenced standards avoid comparisons that can promote an unhealthy competitive atmosphere. Still, many institutions believe a "Norm-referenced" (comparison-based) framework provides a better basis for making judgments about teaching effectiveness. Your report compares your average ratings to results for three different groups of classes. The first comparison group is with all classes in the standard *IDEA database*, and is always reported. The other two are reported only if enough classes were available to provide a stable basis for comparison. These consist of (1) all classes in the *same discipline* as the class in question and (2) all classes *at your institution*. *Institutional* and *disciplinary* norms are updated annually and include the most recent five years of data; the IDEA database is updated on a periodical basis.

➢ More on Criterion Referenced Standards

➢ More on Description of Norms

➢ More on Technical Considerations

Question 3. Were you more successful in facilitating progress on some class objectives than on others? (Refer to the upper portion of Page 2 of the *IDEA Short Form Report*.)

The first portion of Page 2 lists the 12 objectives included on the IDEA form and summarizes student ratings on those you selected as either *Important* or *Essential*. A review of the specific objectives can help you determine where you might focus improvement efforts.

➢ More on Improving Teaching and Learning

The reporting format is similar to that used on Page 1. In addition to "raw" and "adjusted" scores, the report shows the percent of students making ratings in the two lowest categories (No apparent progress or Slight progress) and in the two highest categories (Substantial progress and Exceptional progress). "Converted scores" are shown in the right hand section and compared with the three norm groups previously described (IDEA Database and, if available, Discipline and Institution). In addition to the actual converted average, the report describes the status of each relative to other classes in the comparison group: "Much higher" (highest 10%); "Higher" (next 20%); "Similar" (Middle 40%); "Lower" (Next 20%); or "Much Lower" (lowest 10%). Using broad categories like these rather than precise numbers is a reminder that ratings are neither perfectly reliable nor perfectly valid.

➢ More on Class Objectives

Question 4. Do some salient characteristics of this class and its students have implications for instruction? (Refer to the bottom portion of Page 2 of the IDEA Short Form Report.)

Student Characteristics. Students described their motivation by making self-ratings on the two items listed at the bottom of Page 2. These characteristics have been found to impact student ratings of progress.

➢ More on Impact of Student Characteristics

Page 3 of the *Report* provides a detailed statistical summary of student responses to each of the items on the IDEA form as well as to optional locally devised items, if any.

➢ More on Using Statistical Detail

Additional Information: IDEA Short Form Report

Updated August 2008

Types of Scores

A. Average Scores. "Averages" are simply numerical averages of ratings for the class. All IDEA ratings are made using a 5-point rating scale; but, as described in this *Guide* and on the form itself, the points on the scale have different meanings in different sections of the rating form. You can use the description associated with each point on the rating scale to obtain a verbal characterization of each average.

Although IDEA student ratings have been shown to be both reliable and valid, all measures of human characteristics have some "measurement noise." If students rated the class on another day, results might be somewhat different due to "sampling errors." Such errors are higher for low enrollment classes than for those with large enrollments; for classes in the 15-34 range, a sampling error of ±0.2 is typical. it is slightly higher for smaller classes and lower for larger classes.

One limitation of average scores is that they are higher for some objectives than for others. For example, 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 emphasizing *Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature, etc.)* was 3.69. If only raw averages were considered, instructors choosing the "broad liberal education" objective could be disadvantaged.

B. Converted Averages. One way to compensate for the inequality of average scores on the 12 objectives is to compute "converted" averages. The conversion process results in a group average of 50 and standard deviation (measure of variability) of 10 for each measure. The statistical formula for deriving converted scores is described on page 3 of Technical Report #12. The standard error for converted scores in classes of 15-34 students averages about 3; it is slightly higher for smaller classes and lower for larger classes.

Converted scores make it easier to compare the Progress on Relevant Objective rating for various classes. Those with different "average" (5-point) scores may have the same converted average if they have different objectives. Similarly, "Summary Evaluations" based on converted scores (last column in the table on page 1) are more comparable across classes than are "Summary Evaluations" based on raw or adjusted scores; but these differences are relatively slight.

The chief feature of converted scores is that they supply **normative** comparisons. Numeric scores in the graph compare your ratings with those of all classes in the IDEA database. A score of 50 is "average;" a score of 63 is in the upper 10% of all classes, while one of 37 is in the lowest 10%. If a large enough group of classes was available to ensure stable results, comparisons with other classes in your *discipline* and with other classes at your *institution* are also reported as "converted averages" at the bottom of page 1 and on page 2.

C. Adjusted Ratings. "Extraneous" factors over which the instructor has no control influence student ratings. Adjusted ratings take some of these factors into account. A description of the three factors used to make adjustments on the Short Form is given below.

1. **Student motivation** (Average response to the item, *I really wanted to take this course regardless of who taught it).* Students who respond positively to this item tend to make favorable ratings on items related to course outcomes. Low ratings on this item are an

indication that it is desirable to devote substantial time and effort to improving student interest and involvement before substantive objectives can be successfully addressed. Ratings on this item are a **Major** factor in making adjustments.

2. **Student work habits** (Average response to the item, *As a rule, I put forth more effort than other students on academic work*). Positive responses to this item are related to above average ratings on items related to course outcomes. This is a **Major** factor in making adjustments.
3. **Size of class** (as indicated on the Faculty Information Form). In general, there is a slight tendency for students in large classes to make less favorable ratings than students in small classes. This is a **Minor** factor in making adjustments.

Adjusted ratings are intended to "level the playing field" across classes that differ by purpose, audience, level, size, and types of students. They recognize that conditions beyond the instructor's control can increase or decrease student ratings and, to the degree possible, take these conditions into account by "adjusting" ratings.

Research Report # 6, provides further explanation of the IDEA system extraneous variables.

Criterion Referenced Standards

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. The example provided is not intended for adoption. Institutions electing to establish local "standards" should take into account both the words associated with each point on the rating scale and the consistent tendency for students to make relatively lenient ratings.

An Example of a Criterion Referenced Index

Average Rating	Effectiveness Category
Below 3.0	Below acceptable standards
3.0-3.4	Marginal, improvement needed
3.5-3.9	Good
4.0-4.4	Excellent
4.5 or higher	Outstanding

Criterion-referenced standards are frequently employed when judging a faculty member's qualifications for tenure or promotion and in determining the priority to be given to teaching improvement efforts during the coming year.

Description of Norms

Reliability estimates of ratings in classes with fewer than 10 respondents are too low to permit dependable conclusions; therefore, they were excluded from all norm groups. The IDEA database includes all classes processed between September 1, 1998 and August 31, 2001; all regions of the country; all types of institutions; all levels of instruction; and all disciplines are included. The database includes approximately 45,000 classes, so these norms are highly stable. Norms for the *discipline* and for the *institution* are available only if at least 400 classes were processed during the most recent 5-year period. Norms for progress ratings on individual objectives (page 2 of the report) are available only if the objective was considered *Important* or *Essential* in at least 100 classes.

Technical Report #12 shows that, on an overall basis, there are only slight differences in ratings obtained at various types and sizes of institutions. However, results at a given institution may differ significantly from those obtained at other institutions of the same type. Hence the findings for the "IDEA database" and "Institution" norm groups may differ.

Similarly, Technical Report #13 shows that there are significant disciplinary differences in average ratings. It is uncertain whether this is because some disciplines attract especially effective (or ineffective) teachers or if the inherent characteristics of some disciplines (in their complexity, abstractness, or human interest) invite especially lenient (or harsh) judgments. In the absence of such knowledge, each institution needs to adopt its own policy regarding the emphasis to be given to comparison with each norm group.

Technical Considerations

"Norm-referenced" Results. Even though the vast majority of instructors in higher education are subject matter experts, committed teachers, and experienced professionals, a "norm-referenced" approach necessitates that half will be identified as "below average" for a given norm group. Such a pejorative designation is understandably resented when criterion-referenced ratings meet or exceed pre-established standards. Nonetheless, for various reasons, many institutions need to differentiate among faculty members on the basis of their instructional excellence; norm-referenced ratings address those needs.

Using Converted Averages. Some institutions prefer to use 5-point scale averages on the grounds that the quality of teaching is best judged by the amount of progress students report on the objectives stressed by the instructor; the fact that these ratings are higher for some objectives than for others may simply indicate that teaching is more effective in classes where such objectives are chosen. Those using converted averages argue that instructors choosing objectives where average progress ratings are relatively low should not be penalized for choosing objectives that are particularly challenging. There is no compelling research evidence to support one or the other of these two possibilities.

Adjusted Scores. Special care should be taken in using adjusted ratings in classes where progress and overall ratings were very high (4.2 or above is a recommended cut-off, but each institution needs to carefully consider this and determine the most appropriate level). In these classes, adjusted ratings will almost always be well below unadjusted ratings, not because the instructor was less effective than suggested by unadjusted ratings, but because extraneous factors played such a large role in fostering student achievement that the teacher's opportunity to influence progress was reduced.

Using Categories. Either "criterion-referenced" or "normative" measures are best classified into 3-5 categories defined by a range of scores. This recognizes two important facts: (1) student ratings are neither perfectly reliable nor perfectly valid; (2) students are not qualified to evaluate a number of key aspects of effective instruction. Therefore, The IDEA Center recommends that a comprehensive evaluation process be employed and that student ratings constitute no more than 30-50% of the final judgment.[2]

[2] Suggestions for conducting a comprehensive evaluation of instruction are included in IDEA Paper #36, *Appraising Teaching Effectiveness: Beyond Student Ratings.*

Class Objectives

Knowing the percent of students making ratings in the two highest and two lowest categories is helpful in identifying classes where student outcomes are "bi-modal" (divided fairly evenly between students who profited greatly and those whose sense of progress was disappointing). Bi-modal ratings often occur when a substantial portion of the class lacks the background needed to profit from the course; changes in pre-requisites may be desirable, or you may want to consider the possibility of offering a separate section for those with limited backgrounds. A bi-modal distribution may also reflect differences in preferred learning styles of students; in such instances, you may want to consider presenting material using multiple methods that respond effectively to those with different learning styles.

To understand the nature of bi-modal ratings of progress, it may be helpful to examine the distribution of responses to items 13 -15 (student characteristics; page 2 and 3 of the Short Form Report). Is there evidence of the presence of distinct groups who differ in their motivation, background, or work habits? If so, do these differences have implications for course pre-requisites, for assigning students for group work, or for presenting class material?

It is suggested that you focus first on your most important objectives (those you chose as **Essential**). For each such objective, use the information in the report to judge whether improved outcomes should be a priority. A degree of urgency can be assigned to each objective based on your review of (a) raw and adjusted averages, (b) percent of students rating their progress as "1" or "2," and (c) comparisons with other classes where the objective was selected as "Important" or "Essential." Then apply the same process to objectives chosen as **Important**.

This process of identifying "target" objectives is a useful first step in developing an improvement strategy.

Research has shown that the number of objectives chosen is inversely related to progress ratings. The IDEA Center encourages faculty members to choose only three to five objectives as *Important* or *Essential*; those choosing more than 6 objectives typically receive lower ratings, perhaps because they are trying to do too much or because the objectives chosen were either inappropriate for the course or not meaningfully addressed. If an instructor fails to identify his/her objectives, a rating of *Important* is assigned to all 12 objectives; this usually results in an unrealistic reduction in overall effectiveness ratings (see Research Note #3).

In reviewing progress ratings on individual objectives, many faculty members are stimulated to reconsider their selection of objectives. Sometimes, disappointing progress ratings can be explained by a discrepancy between the instructor's rating of importance and the amount and/or kind of emphasis given to the objective in class sessions and activities.

Improving Teaching and Learning

Using data from the Diagnostic Form, The IDEA Center has conducted substantial research to identify teaching methods that facilitate learning on each of the twelve learning objectives. If you are disappointed in your progress on any of the relevant objectives in your course (identified on page 2 of the Short Form Report), you might want to consider using the Diagnostic Form the next time you seek student input. In the meantime, a number of resources are available.

- POD-IDEA Center Notes were written in cooperation with the Professional and Organizational Development (POD) Network in Higher Education. These brief papers provide detailed suggestions for improving twenty specific teaching methods; references to relevant professional literature are cited for each method.

- POD - IDEA Center ***Learning*** Notes are papers providing background, helpful hints, and additional resources for each of the 12 IDEA objectives.

- Relationship of Teaching Methods to Learning Objectives shows which of the 20 methods are most highly related to progress each of the twelve objectives.

- Research Report # 4 identifies teaching styles that facilitate student progress on specific learning objectives.

- Research Report # 1 focuses on the relationship between various instructional approaches, on one hand, and both instructional objectives and instructional outcomes on the other.

In addition, The IDEA Center continues to conduct an active research program designed to learn more about how course characteristics and outcomes are related. One of these studies examined differences between classes stressing mathematical/quantitative background and other classes (see Research Report #3). Others have shown the impact of factors such as the instructor's previous experience in teaching the course, the instructor's judgment of the adequacy of students' backgrounds, and the degree to which the course emphasized group work, critical thinking, or writing (Research Report #2). Future studies will focus on questions related to whether teaching techniques most closely related to progress differ for classes that are lecture-oriented as opposed to those that emphasize other teaching approaches (collaborative learning, distance education, etc).

Impact of Student Characteristics

"Extraneous" factors over which the instructor has no control influence student ratings. **Adjusted ratings** take some of these factors into account. A description of the three factors used to make adjustments on the Short Form is given below.

1. **Student motivation** (Average response to the item, *I really wanted to take this course regardless of who taught it).* Students who respond positively to this item tend to make favorable ratings on items related to course outcomes. Low ratings on this item are an indication that it is desirable to devote substantial time and effort to improving student interest and involvement before substantive objectives can be successfully addressed. Ratings on this item are a **Major** factor in making adjustments.

2. **Student work habits** (Average response to the item, *As a rule, I put forth more effort than other students on academic work*). Positive responses to this item are related to above average ratings on items related to course outcomes. This is a **Major** factor in making adjustments.

3. **Size of class** (as indicated on the Faculty Information Form). In general, there is a slight tendency for students in large classes to make less favorable ratings than students in small classes. This is a **Minor** factor in making adjustments.

Adjusted ratings are intended to "level the playing field" across classes that differ by purpose, audience, level, size, and types of students. They recognize that conditions beyond the instructor's control can increase or decrease student ratings and, to the degree possible, take these conditions into account by "adjusting" ratings.

Using Statistical Detail

For each item, the distribution of responses (number of students choosing each alternative), the average rating, and the standard deviation of ratings (a measure of variability) are provided. Faculty members are not expected to achieve high ratings on every item. Attention should be concentrated on objectives (items 1-12) chosen as Important or Essential. High ratings on Items 16-18 are also regarded as favorable. For the other items (13-15), averages are descriptive of the course's students but are not useful in making evaluative judgments. Their relevance depends on the nature of the class (its objectives; available learning opportunities; etc.).

Standard deviations of about 0.7 are typical. When these values exceed 1.2, the class exhibits unusual diversity. Especially in such cases, it is suggested that the distribution of responses be examined closely, primarily to detect tendencies toward a bimodal distribution (one in which class members are about equally divided between the "high" and "low" end of the scale, with few "in-between." Bimodal distributions suggest that the class contains two types of students who are so distinctive that what "works" for one group will not for the other. For example, one group may have an appropriate background for the course while the other may be under-prepared; or one group may learn most easily through "reading/writing" exercises while another may learn more through activities requiring motor performance. In any event, detailed examination of individual items can suggest possible changes in pre-requisites, sectioning, or versatility in instructional methods