



VALIDITY OF THE IDEA STUDENT RATINGS OF INSTRUCTION SYSTEM: AN UPDATE

Donald P. Hoyt and Subashan Perera
IDEA Center

Student ratings of instruction have been a major feature of the higher education landscape for the last three decades. Although there continues to be some controversy over the use and interpretation of these ratings, they are by far the most common source used in the evaluation of teaching. Peter Seldin reported that over 88% of the Liberal Arts Colleges he surveyed always use student ratings; no other source was used by more than 70% of these colleges ("How Colleges Evaluate Teaching: 1988 vs. 1998," *AAHE Bulletin*, 50, 1-7), and, in a broader survey by the U.S. Department of Education (cited in the Winter, 1991 issue of *The Department Chair: A Newsletter for Academic Administrators*), 97% of all department heads reported that they used student evaluations to assess the teaching performance of full-time faculty. The evaluation of faculty performance is critically important, affecting not only the lives of individuals but the vitality and excellence of the institutions in which they are employed. Therefore, it is essential that the information used in making these evaluations be valid (i. e., that it assess what it purports to assess).

A Brief Review of Literature on the Validity of Student Ratings

There is substantial literature on the validity of student ratings. Two of the most helpful reviews are those by William E. Cashin: *Student Ratings of Teaching: A Summary of the Research*, IDEA Paper No. 20, Center for Faculty Evaluation and Development, 1988; and *Student Ratings of Teaching: The Research Revisited*, IDEA Paper #32, Center for Faculty Evaluation and Development, 1995. The vast majority of studies have found student ratings to have substantial, but far from perfect, validity (regardless of the standards against which it is assessed). At the same time, there are important aspects of teaching effectiveness which students are unable to evaluate. Peter Seldin explores these factors (*Changing Practices in Evaluation Teaching*, Bolton, MA: Anker, 1999); and IDEA Paper #36 (*Appraising Teaching Effectiveness: Beyond Student Ratings* by Donald P. Hoyt and William H. Pallett, IDEA Center, 1999) offers some practical suggestions for ways to assess them.

Because the IDEA system relies upon students' ratings of progress on objectives chosen by the instructor, it is especially important to document the validity of student self-ratings of achievement. Comprehensive reviews of such studies have been reported by Peter A. Cohen ("Effectiveness of student-rating feedback for improving college instruction: A meta-

analysis of findings," *Research in Higher Education*, 13, 321-341, 1980; "Student ratings of instruction and student achievement: A meta-analysis of multi-section validity studies," *Review of Educational Research*, 51, 281-309, 1981; "An updated and expanded meta-analysis of multi-section student rating validity studies," a paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, 1986), and by Kenneth A. Feldman ("The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies," *Research in Higher Education*, 30, 583-645, 1989). These reviews of dozens of studies offer strong support to the assumption that students can (and do) make acceptably valid estimates of their learning. Of course, there are many practical reasons why student self-estimates cannot be substituted for the instructor's appraisal; but the evidence suggests that the average of these estimates bears a significant relationship to appraisals provided by outside, objective sources.

Studies of the Validity of the IDEA System

1. Students' ratings of their own learning. Several studies of the validity of students' ratings of progress on instructor-chosen objectives have been conducted by the IDEA Center staff. These studies, first reported by Donald P. Hoyt in "Measurement of Instructional Effectiveness," *Research in Higher Education*, 1, 367-378, 1973, have correlated instructor's ratings of the importance of each objective with students' ratings of their progress on those objectives. These studies make three assumptions:

1. Instruction, in general, is at least minimally effective.
2. Instructors make meaningful judgments when they rate the importance of each objective.
3. Students' ratings of their progress on each objective are at least minimally valid.

If all three assumptions are tenable, then there should be a positive correlation between the instructor's ratings of "importance" and the students' rating of "progress."

This has consistently been found to be the case. In the most extensive of these studies, involving over 100,000 classes, the average correlation between these two ratings on the same objective was .252. In contrast, the average correlation between "unrelated" ratings (the instructor's rating on

one objective and the students' rating on another objective) was .019.

This study was recently updated for the new (revised) IDEA form, using data from approximately 8,000 classes taught in the 1998-1999 school year. This study excluded three objectives no longer included on the IDEA student rating form and added five objectives not previously studied. The average of the critical correlations (those between instructor's ratings of the importance of an objective and students' ratings of progress on that objective) was .278, while that for "unrelated" ratings was .017.

It should be noted that the key correlations will be negatively affected if any of the three assumptions is erroneous. It seems safe to say that none of them is likely to be totally true—in some classes, instruction is not even minimally effective; not all instructors are conscientious and thoughtful in their selection of objectives from the list provided by IDEA; and not all students are willing and able to make thoughtful and conscientious ratings of their progress on each objective. Therefore, average correlations of .25 or higher are regarded as firm evidence of the validity of students' ratings of progress. At the same time, it is also evident that such ratings should be considered as only one piece of evidence with respect to overall teaching effectiveness.

2. The relationship between descriptions of teacher behavior and students' ratings of progress. From its inception, the IDEA system has made a careful study of the relationship between the instructor's classroom activities and the ratings students make of their progress on instructor-chosen objectives. A consistent set of conclusions has emerged from the study of over 100,000 classes assessed by the original IDEA form, and confirmed by recent studies of the revised IDEA form.

- a. Significant relationships exist between students' ratings of their learning and of specific classroom approaches; teaching style affects student learning.
- b. The set of instructor behaviors related to progress ratings is different, in sensible ways, for each of the 12 objectives included in the IDEA system¹. Such differentiated results make it clear that students are discerning in their rating of instructors.
- c. Teaching methods related to progress on a given objective vary depending on the size of class; what is effective in small classes isn't always effective in large classes.
- d. For the most part, relationships between what the teacher does and what the student learns are intuitively sensible.

The fact that the findings make intuitive sense is regarded as an important indirect validation of the ratings. As examples, consider the following. The items most closely related to student gains on "Factual Knowledge" are: *The instructor made it clear how each topic fit into the course*; *The instructor explained course material clearly and concisely*; and *The instructor gave tests, projects, etc. that covered the most important point of the course*. Similarly, methods most closely

associated with progress on "Creative Capacities" are: *The instructor changed approaches to meet new situations*; *The instructor explained the reasons for criticisms of students' academic performance*; *The instructor introduced stimulating ideas about the subject*; and *The instructor gave projects, tests, or assignments that required original or creative thinking*. For the objective of "Acquiring Team Skills", the most relevant methods were: *The instructor formed "teams" or "discussion groups" to facilitate learning*; *The instructor involved students in "hands on" projects such as research, case studies, or "real life" activities*; and *The instructor asked students to help each other understand ideas and concepts*. This type of sensible correspondence between what the teacher does and what the student learns is apparent throughout the 12 objectives.

3. Correspondence between independently obtained teacher and student responses. In the revised IDEA system, participating instructors are asked a number of questions which describe important features of their instructional environment. Some of these can be used to test the validity of the system. By assuming that *instructors provided thoughtful and honest answers* and that *student responses are valid*, we can predict how selected responses of instructors will be related to students' ratings of outcomes.

- a. Instructors indicated the degree to which each of seven instructional options was emphasized in their class: writing; oral communication; computer applications; group work; mathematical/quantitative work; critical thinking; and creative/artistic/design endeavor. The following relationships were expected to hold:
 1. In classes where the instructor chose *Communication Skills* as an important or essential objective, students' progress ratings should be higher when "Writing" is given much emphasis than when it is given no emphasis. This assumption was confirmed. In the "No emphasis" classes, average progress rating was 3.33; in the "Much emphasis" classes, this figure was 4.01. The difference was a full standard deviation.
 2. In classes where the instructor chose *Team Skills* as an important or essential objective, students' progress ratings should be higher when "Group work" is given much emphasis than when it is given no emphasis. This assumption was also confirmed. Progress ratings on *Team Skills* averaged 3.51 for the "No emphasis" classes and 4.22 for the "Much emphasis" classes. The difference was 1.2 standard deviations.
 3. In classes where the instructor chose *Creative Capacities* as an important or essential objective, students' progress ratings should be higher when "Creative/artistic/design endeavor" is emphasized than when it is given no emphasis. Again, the assumption was confirmed. Progress ratings on *Creative Capacities* averaged 3.51 for the "No emphasis" classes and 4.23 for the "Much emphasis" group. This difference was more than one standard deviation.
 4. In classes where the instructor chose *Critical Analysis* as an important or essential objective, students' progress ratings should be higher when "Critical thinking" is emphasized than when it is not. This final assumption was also confirmed. Progress ratings on

¹An exception is the similarity in methods most effective in promoting student progress on *Gaining factual knowledge* and on *Learning fundamental principles, generalizations, or theories*.

Critical Analysis averaged 3.60 for the "No emphasis" group and 3.92 for the "Much emphasis" group. This difference was 0.55 of a standard deviation.

Such a marked correspondence between independently made students' ratings and instructor's ratings offers convincing evidence of the validity of both.

- b. In the revised system, instructors indicated the degree to which various circumstances had a "positive" or "negative" impact on student learning. *If these ratings and students' ratings of outcomes were done validly*, there should be a relationship between the two. That this was so is shown by the following results comparing the IDEA system's four global ratings of teaching effectiveness with instructor's descriptions of circumstances which may impact student learning.

Condition	Student Rating	Instructor Rating		
		Positive	Neutral	Negative
Previous experience teaching course.	Progress, rel. obj.	53.6	51.6	50.0
	Increased interest	3.97	3.77	3.67
	Excellent teacher	4.26	4.02	3.94
	Excellent course	3.98	3.72	3.57
Desire to teach course.	Progress, rel. obj.	53.5	51.6	49.3
	Increased interest	3.96	3.73	3.66
	Excellent teacher	4.25	4.11	4.03
	Excellent course	3.97	3.71	3.56
Adequacy of student background.	Progress, rel. obj.	53.9	53.4	51.8
	Increased interest	4.00	3.96	3.77
	Excellent teacher	4.25	4.21	4.08
	Excellent course	3.97	3.91	3.79
Student enthusiasm.	Progress, rel. obj.	54.1	52.9	51.3
	Increased interest	4.05	3.89	3.65
	Excellent teacher	4.32	4.17	3.97
	Excellent course	4.05	3.89	3.66

While these differences are not large, they are all statistically significant ($P < .001$) and all in the expected direction. They add to the body of evidence that the IDEA system provides valid results.

- c. Participating faculty members also identified the principal type of student enrolling in the class—(1) Freshmen/Sophomores seeking to meet "general education" or distribution requirements; (2) Juniors/Seniors seeking to meet "general education" or distribution requirements; (3) Freshmen/Sophomores seeking to develop background needed for their intended specialization; (4) Juniors/Seniors seeking competence or expertise in their academic/professional specialty; and (5) Graduate or Professional School students. If the IDEA system is valid, then classes in the first two categories should stress the objective of *Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature, etc.)* more frequently than classes in the last three categories; students' ratings of progress on this objective should also be higher for classes dominated by those seeking to meet "general education" or distribution requirements. The opposite findings should be obtained on the objective of *Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course.*

All of these expectations were confirmed. The broad liberal education objective was selected as "important" or "essential" in 37% of classes enrolling students seeking to meet general education/distribution require-

ments but only 14% of classes dominated by students seeking specialized education. Furthermore, student ratings of progress on this objective were significantly higher in classes oriented toward students seeking to meet general education/distribution requirements (3.73 vs. 3.44 for specialty-oriented classes). In contrast, the professional preparation objective was stressed in 71% of classes enrolling students seeking specialized education compared to 29% of classes enrolling students seeking to meet general education requirements. And average progress ratings were also much higher for the three groups enrolling specialty-oriented students (4.16 vs. 3.85 for general education classes). These findings add support to the validity of the IDEA system.

- d. Finally, a number of studies have shown that, in general, students give a much higher priority to "preparation for professions" as a college goal than to "gaining a liberal or general education." For that reason, we believed that students enrolling in courses which addressed their professional interests would, in general, be better motivated than those enrolling in courses which addressed less vital concerns (such as general education). The IDEA system attempts to take into account the influence of extraneous circumstances, of which student motivation is a principal component, by providing "Adjusted" ratings. If our assumption is correct about

the advantage which professionally-oriented courses have in terms of student motivation, and *if the IDEA system makes valid corrections for extraneous circumstances*, then "Adjusted" ratings should be higher than "Raw" ratings for classes enrolling students seeking to meet general education/distribution requirements; the opposite should occur for professionally-oriented classes.

In our most recent "validation" study, two "Adjusted" scores were computed. The first (Adj. 1) was based on the standard items on the IDEA form. The second (Adj. 2) added IDEA's "experimental" items to the standard items; therefore, it was expected to be a more precise measure.

"Adjusted" scores were subtracted from "Raw" scores. If the IDEA system is valid, positive differences should be found for classes where student motivation was low (i. e., general/liberal education classes), while negative differences should be found for classes where student motivation was high (i. e., specialization classes). The findings for each of the five types of classes are shown below.

Type of student enrolled	Adj. 1	Adj. 2
Fresh/Soph—general education	+6.9	+9.3
Jr/Sr—general education	+6.9	+6.3
Fresh/Soph—specialized	-1.7	-1.9
Jr/Sr—specialized	-4.4	-6.5
Grad/Professional	-5.9	-10.2

These data further confirm the general validity of the IDEA system.

Conclusions

Although the professional literature strongly supports the general validity of student ratings of instruction, it is important to document the validity of specific student rating programs. This is especially important for the IDEA system, since its reliance on student estimates of their progress on instructor-chosen objectives makes it unique.

Studies conducted on results from about 8,000 classes in which the revised IDEA form was administered in 1998-1999 added significantly to the documented validity of the IDEA system. First, this study verified previous conclusions that there was a significant relationship between the instructor's ratings of the importance of a given objective and students' ratings of their progress on that objective. Such a finding could occur only if instruction, in general, is at least minimally effective, instructors make meaningful ratings of the relevance of various objectives, and students make at least minimally valid ratings of their progress on these objectives.

The most recent study also reaffirmed previous findings that

the teacher's classroom behavior and students' ratings of progress were related in a different, and logically sensible, fashion for each of the 12 learning objectives. Such differentiated findings could not occur unless students were relatively discerning in their descriptions of instructors and their ratings of progress on learning objectives. This discernment was further revealed by showing that teaching behaviors most closely related to a given type of progress were intuitively logical.

New kinds of validity evidence were made possible by changes made in the *Faculty Information Form* used with the revised IDEA form. Faculty ratings on this form, whether of classroom emphases (e. g., "Writing", "Group Work", etc.), the environment of the class (e. g., the faculty member's desire to teach the class; adequacy of student background; etc.), or the intended audience (e. g., lower division students seeking to satisfy general education/distribution requirements; upper division students seeking expertise in their specialty), were related to independently obtained student ratings of outcomes in ways which would be expected if the IDEA system were valid.

With these studies, it seems safe to conclude that the IDEA system has a highly acceptable level of validity for its purposes. It may well be the most extensively validated student rating program in the nation.

IDEA Center
215 Seth Childs Road, Suite B
Manhattan, KS 66502
Tel: 800.255.2757 or 785.532.5970
E-mail: IDEA@ksu.edu
Home page: www.idea.ksu.edu

©2000, IDEA Center