PRECIS

REVISING THE IDEA SYSTEM FOR OBTAINING STUDENT RATINGS OF INSTRUCTORS AND COURSES

Donald P. Hoyt, Yih-Fen Chen, William H. Pallett, and Amy B. Gross
The IDEA Center, Kansas State University

What follows is a detailed explanation of processes employed, and results obtained, in preparing a revision of the IDEA system. The report is lengthy and includes 30 tables which summarize the statistical analyses which were conducted to guide the revision process.

Although a few readers will have a keen interest in the detailed figures, others will be interested in a more macroscopic view of the project or some of its parts. For such readers, this *Precis* has been prepared. It poses eight questions and describes where the answers to these questions are found in the full report.

Question 1. In capsule form, what did you do and what did you find.

A summary of the project and its findings is provided on pp. 52-53.

Question 2. Why was a revision needed?

The **Background** provided in Section I, p. 1, answers this question.

Question 3. Were IDEA's most distinguishing features retained?

Yes. See Part II, **The Revision Strategy**, on p. 1.

Question 4. How were potential new items developed?

See Part III, **The Revision Process: Identifying Potential Additional Items**, pp. 2-4. This section describes how both potential new objectives and potential new teaching methods items were identified.

Question 5. How were items for the new form selected?

Part IV, **Selecting Items for the Revised Form**, responds to this question; see pp. 4-39. A major strategy required the used of a Pilot Study. The design for this study is reported on pp. 4-5; it's use in selecting objectives for the revised form is described on pp. 5-7.

The rationale for selecting items is reviewed on pp. 7-8. Data from both the Pilot Study and IDEA's national data base were used to discover which items made the most contribution to the prediction of student ratings of progress on a given objective. Four types of analyses were required:

1. The utility of existing methods items for predicting progress on existing objectives (pp. 8-16).
2. The utility of existing methods items for predicting progress on proposed objectives (pp 17-20)
3. The utility of proposed items for predicting progress on existing objectives (pp. 20-28)
4. The utility of proposed methods items for predicting progress on proposed objectives (pp. 28-

35).

In addition to the empirical analyses reported on pp. 8-35, consideration was given to ensuring (a) that important dimensions of instructional effectiveness were represented (pp. 36-37) and (b) that the instrument's diagnostic information would be as comprehensive as possible (pp. 37-39).

Question 6. How were outcome measures "adjusted" to take into account extraneous influences (factors beyond the instructor's control)?

The problem is discussed on pp. 39-40. The specific extraneous measures considered are described on pp. 40-41. The type and degree of influence they have on outcome measures is reviewed on pp. 41-44.

Question 7. How reliable are the various measures used by the IDEA System?

Reliability information for individual items is provided on pp 44-47. Comparable figures for more complex measures (combinations of items) are given on pp. 47-48.

Question 8. How are teaching "strengths" and "weaknesses" identified?

Both class size and student "motivation level" are considered, as explained on p. 48. Empirical findings are reported on pp. 49-51.

REVISING THE IDEA SYSTEM FOR OBTAINING STUDENT RATINGS OF INSTRUCTORS AND COURSES

Donald P. Hoyt, Yih-Fen Chen, William H. Pallett, and Amy B. Gross
The IDEA Center, Kansas State University

I. Background

The IDEA system was initially developed by Hoyt (1973) in 1969. Based on feedback from users, measurement and instructional authorities, and colleagues, a number of revisions and refinements were made over the next few years. The last of these was made when the Center for Faculty Evaluation and Development was established, with support from the Kellogg Foundation, in 1975. For the next 23 years, the form for collecting student ratings has been basically unchanged; and only minor changes have been made in the way the results were summarized and reported.

From 1975 through 1996, the Center has had three directors--Dr. Bert Biles, Dr. Lance Kramer, and Dr. William Cashin. Through their leadership, it created the tools needed to develop a timely and reliable system of providing results to users. Supporting materials were also published as a national database was established. And supplementary aids to instructional improvement, including the *IDEA Papers, Technical Reports*, and a series of national seminars on related topics were offered.

Dr. William Pallett became the director of the Center when Dr. Cashin retired in January, 1997. Dr. Hoyt, who retired in 1992, accepted a part time appointment to direct the research effort for the Center. In the early part of 1997, Pallett and Hoyt committed the Center to developing the first revision of the IDEA system in 23 years.

The need for a revision was obvious. Between 1975 and 1998, there have been many changes in higher education. The widespread development and accessibility of educational technology is a dramatic example. There have been many other changes also--in instructional approaches, in educational purposes and objectives, in teaching-learning conceptualizations, and in the settings in which instruction is offered. Given these changes, it seemed apparent that an instrument developed in the early 1970s could not be highly responsive to the needs of a new millennium.

II. The Revision Strategy

In developing a strategy for revision, a commitment was made not to alter materially the length of the forms and to retain the features of the IDEA system which distinguishes it from most others:

1. It would continue to focus on <u>student learning</u> as the chief indicator of teaching effectiveness.
2. It would continue to take into account factors which affect student ratings but which are not under the instructor's control.
3. It would continue to focus on guiding instructional <u>improvement</u> efforts.

Revision decisions were to be made by the IDEA professional staff. They were to be guided by suggestions from users, the professional literature, and authorities in the field. Insofar as possible, changes were to be consistent with empirical research findings.

III. The Revision Process: Identifying Potential Additional Items

The initial step in revising the form was a survey of users conducted in the spring of 1997. Slightly different survey forms were used for those who employed the long (diagnostic) and short forms[1].

Although the surveys asked questions about IDEA services and the support provided for them, the key questions for the purposes of this report revolved around suggestions for new learning objectives, for new teaching methods, for descriptions of the course and its enrollees, and for making the report of results most useful.

Replies to the "long form" survey were received from 32 institutions; multiple users from 7 of these provided independent responses. Several of these institutions also used the short form, and some of their replies were relevant to both forms. Another 13 institutions responded to the short form survey only.

In general, respondents agreed that the report format needed extensive attention. Most reported that faculty found it so complex and technical that it was difficult to understand. About half thought that the amount of detail was excessive, but almost as many thought that detailed reporting was essential to accurate understanding and interpretation. In any event, it was clear that a major overhaul of the report was desired[2].

Objectives. The survey form used "Learning how to learn" and "Learning to be an effective member of a team" as examples of objectives which reflect contemporary educational thought but which were not represented on the IDEA forms. Most respondents agreed that these were important learning objectives. A number of other suggestions were also made. Although none of these was identified by a majority of respondents, an attempt was made to include paraphrases of them, along with the two identified above, in a preliminary list of potential new objectives. Through staff discussion, six such objectives were formulated:

1. Acquiring skills in working with others as a member of a team
2. Learning how to find and use resources for answering questions or solving problems on my own
3. Understanding the relationships between this field of study and other disciplines.
4. Acquiring increased intellectual curiosity (the disposition to learn more by asking questions and seeking answers).
5. Learning to analyze and critically evaluate ideas, arguments, and points of view.
6. Developing a clearer understanding of, and commitment to, values which will guide the development of personal practices and life style.

---

[1] Copies of the surveys are available upon request from the IDEA Center.

[2] The new format features a logical organization (outcomes first, then methods, and finally a description of the teaching environment), graphic presentations of results, and interpretive aids as an integral part of the report.

A select group of IDEA users with expertise in teaching evaluation and faculty development[3] were asked to review and critique this list. On the basis of their feedback, some minor changes were made on three of these:

2. Learning how to find and use resources for answering questions or solving problems.
4. Acquiring an interest in learning more by asking my own questions and seeking answers.
6. Developing a clearer understanding of, and commitment to, personal values.

In addition, because there was some confusion about the distinction between the fifth proposed objective (Learning to analyze and critically evaluate ideas, arguments, and points of view) and one of the original objectives [Learning to apply course material (to improve thinking, problem solving, and decisions)], it was decided to italicize key words which differentiated them. The italicized words were *analyze and critically evaluate* (proposed objective 5) and *apply* (original objective).

These six objectives, as revised, were included in the pilot student reported below.

Teaching Methods. Survey respondents identified a number of teaching methods which they regarded as effective but which were not covered by the current form. These included items related to technology, assessment, applications, cooperative learning, problem-based learning, and "hands-on" instruction. The most common suggestion was that the form should provide coverage of the Seven Principles for Good Practice in Undergraduate Education (Chickering and Gamson, 1987). Accordingly, an attempt was made to construct new "methods" items which included the specific suggestions of our reviewers but which organized these around the seven principles.

A set of 26 new items was drawn up. The staff agreed that no more than 15 could realistically be included in a pilot study. After considerable discussion, the following items were proposed:

1. Used educational technology to clarify course concepts and/or expand understanding.
2. Provided ample opportunity to obtain assistance outside of class.
3. Encouraged students to work together on projects, class presentations, or other assignments.
4. Relied on a variety of assessment methods, not just tests, to evaluate student achievement and award grades.

[3]We are grateful from the thoughtful suggestions offered by V.Clegg, Kansas State University; D. Fink, University of Oklahoma; L. Sorenson, Brigham Young University; P. O'Brien, Bridgewater State University; R. Sherry, Bethel College; B. J. Johnson, University of San Francisco; K. Carey, Eastern Kentucky University; and B. Walvoord, Notre Dame University.

5. Scheduled homework assignments, tests, and projects in ways which encouraged students to stay up to date in their work.
6. Encouraged students to use multiple resources (e. g., data banks, library holdings, computer programs) to improve understanding.
7. Encouraged students to help each other understand ideas or concepts.
8. Shared relevant personal experiences, attitudes, and values with class.
9. Required that students complete course assignments according to schedule.
10. Provided prompt feedback on tests, reports, projects, etc.
11. Encouraged student-faculty interaction outside of class.
12. Established high expectations in this class.
13. Encouraged discussion/debate among those with diverse backgrounds.
14. Did things that discouraged students from postponing work on the course.
15. Placed responsibility for learning squarely on the students.

These items were submitted to the same experts who provided feedback on the proposed new objectives (footnote #2). They suggested a number of additions, deletions, and editorial changes. Only Item 8 from the above list was left intact. Items 1, 2, 3, and 9 were deleted and replaced with:

1. Displayed a personal interest in me and my learning.
2. Formed "teams" or "discussion groups" to facilitate learning.
3. Involved students in "hands on" projects such as research, case studies, or "real life" activities.
9. Inspired students to set and achieve goals which really challenged them.

Editorial changes were made in the other 10 items, resulting in the following wording:

4. Relied on a variety of assessment methods, not only tests, to evaluate student progress on course objectives.
5. Scheduled course work (class activities, tests, projects) in ways which encouraged students to stay up to date.
6. Encouraged students to use multiple resources (e. g., data banks, library holdings, outside experts) to improve understanding.
7. Asked students to help each other understand ideas or concepts.
10. Provided timely and frequent feedback on tests, reports, projects, etc. to help students improve
11. Encouraged student-faculty interaction outside of class (office visits, phone calls, e-mail, etc.)
12. Established high expectations of students in this class.
13. Asked students to share ideas and experiences with others whose backgrounds and

viewpoints differ from their own.

14. Created incentives that facilitated the timely completion of course assignments.
15. Required students to take their share of responsibility for learning.

With these changes, the list of pilot items (six new objectives; fifteen new teaching methods) was complete. A pilot test designed to provide empirical evidence to aid in the selection of no more than 12 objectives and no more than 20 methods items for the revised IDEA form.

IV. Selecting Items for the Revised Form

The Pilot Test: Design. Six different lists of proposed new items were prepared, each of which contained three new objectives and five new methods. A counter-balancing scheme was employed to ensure that, across the six sets, each objective would be paired with each of the 15 methods. In this way, pilot institutions would be asked to collect ratings on only 8, rather than 21, new items; but information would be collected which related each of the proposed new methods to each of the proposed new objectives (as well as new methods to "old" objectives, and "old" methods to new objectives).

A list of 14 institutions of various types (public universities; private colleges; community colleges) which had regularly participated in the program on a relatively large scale was prepared. The IDEA program coordinator at ten of them[4] agreed to collect responses to one set of pilot items as "extra questions" during the fall, 1997, IDEA administration. Participating faculty at these institutions completed an augmented Faculty Information Form which included ratings of the importance of all six of the proposed new objectives[5].

Before compiling results related to the proposed new items, a cursory study was made to determine the degree to which IDEA results for the ten cooperating institutions were typical of those in the IDEA database. For this purpose, data for seven items of various types were

[4] Central Missouri State University; Tennessee Technological University; Morehead State University (Kentucky); Howard Community College; Truman State University; Wichita State University; Bethel College (Minnesota); Eastern Kentucky University; Johnson County Community College; and Southeast Missouri State University.

[5] Faculty ratings on all six objectives were requested even though students in their classes rated progress on only three of these in order to increase data about the relative "popularity" of these objectives with faculty.

examined--three "methods"items (*4. Seemed enthusiastic about the subject matter; 15. Stimulated students to intellectual effort beyond that required in most courses; and 17. Explained course material clearly, and explanations were to the point.);* one course characteristic (*33. Difficulty of subject matter)*; one student characteristic (*36. I had a strong desire to take this course.)*; and two outcome items (*44. Overall, I rate this instructor an excellent teacher; 46. Overall, I learned a great deal in this course.*). Means for the pilot group were compared with those for the rest of the institutions in the database, with the following results:

Item	Pilot Mean	Database Mean
4	4.50	4.49
15	3.66	3.69
17	4.10	4.11
33	3.28	3.35
36	3.62	3.56
44	4.17	4.18
46	4.10	4.11

Correlations between the two criteria (means scores on Items 44 and 46) and the other five items were also computed for both the pilot group and for the rest of the database, with the following results:

	Correlation with Item 44		Correlation with Item 46	
Item	Pilot	Database	Pilot	Database
4	.79	.76	.65	.65
15	.66	.68	.68	.69
17	.89	.88	.77	.76
33	-.15	-.10	-.02	-.03
36	.36	.39	.57	.60

From these data, it appeared that results from the ten pilot institutions would provide a good proxy for those obtained from the much larger data base.

Materials needed to participate as a "pilot institution" (one set of 8 pilot items; instructions to student raters; instructions to faculty) were prepared by the IDEA Center and mailed to the campus coordinator well in advance of the fall, 1997, program. Of course, participation in the tryout effort was voluntary on the part of the faculty member. Pilot participation rates varied from 45.7% to 92.9% among institutions for all classes surveyed with the IDEA instrument; the average was 67.3%.

Using the Pilot to Select Objectives. A critical concern was the frequency with which objectives (proposed and existing) were chosen as "Important" or "Essential" by participating faculty. Existing objectives were rated in all 3668 tryout classes; the number of classes for which ratings

were made of the proposed objectives varied from 845 to 2052. Results for "Existing" and "Proposed" objectives are given in Table 1.

TABLE 1

Frequency With Which Various Objectives Were Chosen As "Important" or "Essential" in Tryout Classes

Existing Objective	**Number Rated**	**% "Imp" or "Essential"**	**% "Essential"**
1. Gaining factual knowledge.	3668	62.3	38.7
2. Principles, theories, generalizations.	3668	61.2	30.8
3. Applications for thinking, problem solving.	3668	51.3	21.4
4. Professional skills and viewpoints.	3668	44.0	17.1
5. Learning how new knowledge is developed.	3668	18.9	3.4
6. Developing creative capacities.	3668	14.7	5.4
7. Personal responsibility, self-reliance.	3668	27.6	6.2
8. Broad understanding, appreciation, of cultural-intellectual activity.	3668	14.2	5.4
9. Communication skills (written; oral).	3668	32.9	14.5
10. Implications for self-understanding.	3668	19.9	7.7
Proposed Objectives			
1. Acquiring "team" skills.	845	56.0	33.6
2. Finding, using resources for info, problem solving.	1126	19.8	9.5
3. Understanding relationship with other disciplines.	1126	19.9	7.0
4. Understanding, committing to personal values.	1773	36.1	22.5
5. Asking own questions, seeking answers.	1659	25.4	6.4

6. Analyzing, critically evaluating ideas, etc.	2052	38.7	16.4

Based on the frequency of selection, pilot results suggested that at least six of the existing (E) objectives should be retained. In addition, three of the proposed (P) objectives were also chosen in at least 30 percent of the pilot classes. Using this "popularity" standard, the objectives tentatively selected for the revised form were:

1E. Gaining factual knowledge (terminology, classifications, methods, trends).
2E. Learning fundamental principles, generalizations, or theories.
3E. Learning to apply course material to improve rational thinking, problem-solving, and decision making.
4E. Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course.
7E. Developing a sense of personal responsibility (self-reliance, self-discipline).
9E. Developing skill in expressing myself orally or in writing.
1P. Acquiring skills in working with others as a member of a team.
4P. Developing a clearer understanding of, and commitment to, personal values.
6P. Learning to analyze and critically evaluate ideas, arguments, and points of view.

In addition to these nine relatively popular objectives, Objective 6E (Developing creative capacities) was tentatively chosen for inclusion on the basis that, although not chosen often, it was a such a unique objective that it could not easily be represented by a combination of the other objectives. Feedback from some reviewers indicated that this objective may be seldom chosen because it was viewed as relevant only to art and music; therefore, it was decided to augment it with a description of the ways creative capacities might be expressed (writing, inventing, designing, performing in art, music, drama, etc).

Two of the proposed objectives were developed to reflect the focus on “life-long learning” as an objective of higher education (2P. Learning how to find and use resources for answering questions or solving problems; 5P. Acquiring an interest in learning more by asking my own questions and seeking answers). Objective 5P was selected as “important” or “essential” by more faculty members than any of the remaining objectives; and Objective 2P was selected as “essential” more frequently than any other objective. For these reasons, and because “life-long learning” has, in the last decade, been so prominently featured as a learning objective by higher education leaders, these two were tentatively selected as the final objectives to be included on the revised form[6].

Selecting Methods Items: Rationale The chief criterion for selecting the 20 methods items for the revised form was the usefulness of the item in predicting student progress ratings on the learning objectives, especially the 12 objectives tentatively chosen for the revised form. A secondary

[6]Ultimately, Objective 8E [*Gaining a broader understanding of intellectual/cultural activity (music, science, literature, etc.)*] replaced Objective 7E [*Developing a sense of personal responsibility (self-reliance, self-discipline)*] in the revised form. The rationale for this change is given in a subsequent section of this report.

concern was to select items to represent meaningful "dimensions" of teaching styles[7]. As noted above, 35 methods items were considered--20 from the existing form and 15 proposed new items.

[7]The Chickering-Gamson *Seven Principles* (1987) was used as the basis for classifying items into "teaching style dimensions". Through staff discussion, these principles were combined into four (tentative) dimensions: (1) Student-Faculty Contact; (2) Involving Students; (3) Establishing Expectations; and (4) Assessment/Feedback. A fifth dimension, Clarity of Communication, was added.

The IDEA system has traditionally considered a given "Methods" item as relevant if ratings on that item correlated substantially with progress ratings on a given objective. While such correlations are always of interest, they can be somewhat misleading. This is because there is considerable overlap among ratings of these items (Sixbury and Cashin, 1995). For example, ratings on Item 1 (Promoted teacher-student discussions) and those on Item 3 (Encouraged students to express themselves freely and openly) both correlate highly (.62 and .64, respectively) with progress ratings on Item 26 (Developing creative capacities). However, ratings on these two methods items correlated .84 with each other. Once the relevance of Item 3 to progress ratings on Item 26 is taken into account, no additional (unique) information is provided by Item 1[8].

In the present study, we sought to determine the degree to which each of the "methods" items (20 "existing" and 15 "proposed") made an independent contribution to the prediction of each progress rating. While all progress ratings were considered (10 existing; 6 proposed), special attention was given to the 12 which had been tentatively selected for inclusion on the new form.

Data Sources For determining the degree to which the proposed items made unique contributions to the prediction of progress ratings, the only available data were those from the 3668 classes in the pilot institutions. For analyses of existing items, we chose to employ 35,458 classes rated by at least 10 enrollees during the 1994 and 1995 academic years--a more current source of data than the 100,000 class database compiled by Sixbury and Cashin (1995). Thus, two data sources were used:

1. The pilot group used to examine the relationship between (a) the proposed methods items and the proposed objectives, (b) the proposed methods items and the existing objectives, and (c) the existing methods and the proposed objectives.
2. An "historical" (1994-95) database which was used to examine the relationship between the existing methods and the existing objectives.

Using "Existing" Methods Items to Predict "Existing" Objectives. Analyses reported in this section were based on the historical database. The general design required (a) selecting classes for which a given objective was relevant (had been chosen by the instructor as "Important" or "Essential"), (b) sorting such classes into one of four sizes--Small (10-14); Medium (15-34); Large (35-99); and Very Large (100 or more); and (c) conducting step wise multiple regression analyses to determine which of the 22[9] existing methods made a unique contribution to the prediction of the objective in question. Four multiple regression analyses were conducted for

[8]High item inter-correlations occur in part because instructors who are skilled in one approach tend to be skilled at another also, and in part because there is a dynamic dependency among items [i. e., if one is good at promoting teacher-student discussion (Item 1), this may be so because she/he encourages free expression by students (Item 3)]. An additional important explanation for the high degree of item overlap is the "halo effect" (the tendency for raters to form a general impression of the teacher and make all ratings conform to this).

[9]Although the original IDEA form contains only 20 "methods" items, it employs seven "experimental" items, including two on assessment methods ("The instructor gave tests, projects, etc. that covered IMPORTANT POINTS of the course" and "The instructor gave projects, tests, or assignments that required ORIGINAL OR CREATIVE THINKING"). These two items were included in the analyses reported in this section.

each of the ten objectives (one for each size)[10].

In the step wise procedure, the item most closely related to the criterion (rating of progress on a given objectsive) is selected first. Subsequent items are selected in accord with the amount of additional criterion variance "explained" or accounted for. Item selection ceases when the additional variance explained is not statistically significant (P<.01).

A summary of the main findings for the seven objectives which were ultimately retained on the revised form is provided by Tables 2-8[11].
Two of the objectives *Gaining factual knowledge (terminology, classifications, methods, trends)* and *Learning fundamental principles, generalizations, or theories* (Tables 2 and 3) are concerned with
the development of relatively low level cognitive skills/knowledge. Student boredom represents a serious

TABLE 2

Methods Items Selected as Relevant to "Gaining Factual Knowledge" In Small, Medium, Large, and Very Large Classes

Methods Item	Class Size							
	Small		Medium		Large		Very Large	
	b	r	b	r	b	r	b	r
3. Encrgd free/open expression	--	--	-11	54	--	--	--	--
4. Seemed enthusiastic	--	--	--	--	-16	61	-14	64
10.Made it clear how topics fit	26	69	30	74	36	78	22	80
13.Encrgd student comments	--	--	--	--	-12	53	-07	63
15.Stimulated intellectual	25	65	25	68	29	73	32	74

[10]Multiple regression analysis requires stable estimates of all correlations. Therefore, it was not undertaken if there were fewer than 100 classes in any given size.

[11]The IDEA Center will provide detailed results on the other three objectives upon request.

effort								
20.Introduced stimulating ideas	18	67	21	68	19	73	17	75
40.Tests covered important pts.	25	60	30	66	29	72	46	78
41.Assess. required originality	-08	40	-08	37	--	--	--	--
N	4808		17,520		4646		218	
R^2	.604		.664		.716		.773	

Note: Decimals associated with "b" and "r" have been omitted.

threat to student progress on such goals which may explain the relevance of Item 15 for both objectives,
regardless of class size. The relevance of Item 20 for the first of these objectives can also be accounted for on this basis. Course organization and integration (Item 10) and ensuring that tests or projects covered the course's most important points (Item 40) were the other two items which consistently made independent contributions to the prediction of progress ratings on these objectives. It appears that, for courses emphasizing straightforward cognitive objectives--those requiring no applications, inferences, or reasoning--success is most dependably promoted by stimulating curiosity and effort, by effective course planning, and by designing fair and comprehensive tests or projects.

Other methods items made small independent contributions to the prediction of progress ratings on these objectives for some class sizes, but not for others. It should be noted that, on a few items, the regression weights were negative, even though their simple correlation (r) with the progress rating was positive. The statistical explanation for this phenomenon is found in the concept of "suppressor variable", a subject which is too complex for extended discussion here (Horst, 1941). Suffice it to say that, in large classes, the

TABLE 3

Methods Items Selected as Relevant to "Learning Fundamental Principles. . ." In Small, Medium, Large, and Very Large Classes

Methods Item	**Class Size**							
	Small		**Medium**		**Large**		**Very Large**	
	b	**r**	**b**	**r**	**b**	**r**	**b**	**r**
2. Helped Ss answer own								

Qs	--	--	--	--	--	--	17	80
4. Seemed enthusiastic	--	--	--	--	-12	60	--	--
7. Spoke with expressiveness	--	--	--	--	--	--	-09	59
10.Made it clear how topics fit	20	68	34	73	22	75	--	--
15.Stimulated intellectual effort	27	66	27	68	36	74	27	76
17.Explained clearly	17	67	--	--	--	--	18	78
18.Related to real life	--	--	--	--	11	61	--	--
40.Tested important points	18	60	25	65	29	71	32	78
N	4763		17,709		4491		199	
R^2	.600		.639		.698		.794	

Note: Decimals associated with "b" and "r" have been omitted.

more the teacher displayed the method in question [e. g., seemed enthusiastic (Item 4) or encouraged student comments (Item 13)], the less likely it is that students would report progress on these cognitive objectives. One can speculate that, in such classes, the positive effect of "showmanship" (displaying enthusiasm) and encouraging student comments was overcome by their negative impact on techniques (intellectual stimulation; tight organization) which effectively promote progress on low level cognitive objectives.

Findings with respect to the objective which focuses on applications *(Learning to apply course material to improve rational thinking, problem solving, and decision making)* are reported in Table 4. There was some overlap between approaches which effectively promoted progress on this objective and those which were effective for the two cognitive objectives just reviewed, particularly in terms of stimulating intellectual effort (Item 15) and testing for the most important points (Item 40). But there were three other techniques which consistently made independent contributions to the prediction of progress ratings on this goal--helping students answer their own questions (Item 2), relating course concepts to real life (Item 18), and designing tests or projects which required original thinking (Item 41). This is clearly a higher level cognitive objective, and progress on it seems to be facilitated by encouraging student responsibility for learning, by helping students to relate content to everyday life, and by demanding that students demonstrate their originality/creativity.

Interestingly, Item 4 ("Seemed enthusiastic about the subject matter") consistently carried a negative weight, even though it was positively correlated with progress ratings. This finding suggests that, although appearing to be enthusiastic (energetic; excited) may, by itself, promote

improvement in students' ability to apply course content, it may also detract from the instructor's ability to use more potent instructional approaches (e. g., offering real life examples; creating student involvement in problem solving). This would explain its overall negative impact on progress ratings.

TABLE 4

Methods Items Selected as Relevant to "Applications for Thinking, Problem Solving" In Small, Medium, Large, and Very Large Classes

Methods Item	Class Size							
	Small		Medium		Large		Very Large	
	b	r	b	r	b	r	b	r
2. Helped Ss answer own Qs	18	68	28	74	39	77	65	79
3. Encouraged free expression	--	--	--	--	--	--	-21	66
4. Seemed enthusiastic	-14	55	-21	59	-23	58	-33	50
8. Demo. import of subject	--	--	18	72	--	--	39	73
10.Made clear how topics fit	18	68	--	--	--	--	--	--
11.Explained criticisms	--	--	--	--	--	--	-20	49
13.Encrgd student comments	--	--	--	--	-15	59	--	--
15.Stimulated intellectual effort	28	68	27	71	31	75	--	--
18.Related to real life	19	62	20	64	28	67	23	65
40.Tested important points	14	57	18	59	21	64	28	62
41.Assess. required originality	08	54	--	--	08	60	14	63
N	4302		16,010		3520		161	
R^2	.631		.679		.739		.799	

Note: Decimals associated with "b" and "r" have been omitted.

The fourth objective, *Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course*, is also an "applications" item, albeit one focused on vocational/professional development rather than on capacities for thinking and problem solving. Findings are reported in Table 5. Both of these objectives were promoted by the instructor's ability to stimulate intellectual effort (Item 15) and to relate the course to real life (Item 18). Except for very large classes (of which there were only 102 in the database), progress ratings on this objective were also improved by attention to course integration (Item 10) and to providing constructive feedback (Item 11).

TABLE 5

Methods Items Selected as Relevant to "Professional Skills, Viewpoints" In Small, Medium, Large, and Very Large Classes

Methods Item	Class Size							
	Small		Medium		Large		Very Large	
	b	r	b	r	b	r	b	r
2. Helped Ss answer own Qs	--	--	--	--	23	74	31	82
4. Seemed enthusiastic	--	--	--	--	-31	60	--	--
6.Exams required rote	--	--	--	--	--	--	-17	-54
8. Demonstrated importance	20	72	--	--	27	77	--	--
10.Made clear how topics fit	21	72	72	76	24	78	--	--
11.Explained criticisms	12	56	14	59	16	65	--	--
13.Encrgd student comments	--	--	--	--	-15	57	-22	63
15.Stimulated intellectual effort	19	64	18	67	24	73	30	76
18.Related to real life	14	63	20	66	21	69	29	64
40.Tested important points	13	58	13	58	--	--	--	--
41.Assess. required originality	--	--	--	--	--	--	16	73
N	4123		12,565		2428		102	

R^2	.631	.665	.743	.796

Note: Decimals associated with "b" and "r" have been omitted.

Class size made a difference on some items. For large and very large classes, responses to Item 2 ("Found ways to help students answer their own questions") made an independent, positive contribution to the prediction of progress ratings, while Item 13 ("Encouraged student comments even when they turned out to be incorrect or irrelevant") made a negative contribution. Although Item 13 was positively correlated with progress ratings, its negative contribution to prediction probably means that, by encouraging student comments, the instructor inadvertently reduced his/her ability to employ more potent methods. For small and medium sized classes, Item 40 ("Gave tests, projects, etc. that covered important points of the course") made a positive contribution to progress ratings.

Item 8 ("Demonstrated the importance and significance of the subject matter") made a substantial independent contribution to the prediction of progress ratings in both small and large classes. Although its partial regression weight was not significant in medium or very large classes, its zero-order correlation was over .75 in both instances. It is reasonable to assume that this method, like Item 18 ("Related course materials to real life situations"), generally promotes professional development by providing an ego-involving context for the subject matter.

As noted earlier, the fifth objective, *Developing creative capacities,* was retained because of its relatively unique focus. As Table 6 shows, there were only a few methods which consistently predicted progress ratings on this criterion. However, the amount of variance accounted for (R^2) was very high.

TABLE 6

Methods Items Selected as Relevant to "Creative Capacities" In Small, Medium, and Large Classes*

Methods Item	Class Size					
	Small		Medium		Large	
	b	r	b	r	b	r
5. Changed approaches for new situations	21	66	14	69	--	--
10.Made it clear how topics fit	--	--	08	63	--	--
11.Explained criticisms	15	60	12	63	15	65
13.Encouraged student comments	--	--	--	--	-15	57

15.Stimulated intellectual effort	--	--	--	--	17	71
17.Explained material clearly	--	--	--	--	22	67
18.Related to real life	-15	46	-16	51	--	--
19.Gave picky exam questions	--	--	--	--	-15	-48
20.Introduced stimulating ideas	37	73	39	75	33	79
40.Tested important points	--	--	--	--	-31	39
41.Assess. required originality	42	73	44	76	36	70
N	1972		6405		854	
R^2	.705		.754		.793	

* N for "Very Large" classes was 46, too low to permit useful analysis.
Note: Decimals associated with "b" and "r" have been omitted.

Three methods made independent contributions to the prediction of progress ratings regardless of class size
(11. "Explained the reasons for criticism of students" academic performance"; 20. "Introduced stimulating ideas about the subject"; and 41. "Gave projects, tests, or assignments that required original or creative thinking"). Apparently, if instructors expand student horizons and provide rational, creatively-oriented assessments and feedback, their students are likely to experience growth in creative capacities. For small and medium sized classes, two other techniques were found to be relevant: flexibility (Item 5--"Changed approaches to meet new situations") and, negatively, real life applications (Item 18--"Related course material to real life situations"). The negative regression coefficient for the latter suggests that, by focusing on the practical, instructors may inhibit growth in creativity.

Except for the findings already reported with respect to Items 11, 20, and 41 (stimulation, feedback, assessment stressing originality), results for large classes (enrollments of 35-99) were relatively distinctive. Gains on this criterion were higher for instructors who stimulated student intellectual effort (Item 15) and who explained clearly (Item 17), but lower for those who encouraged student comments (Item 13), gave examinations containing "picky" questions (Item 19), and focused exams on the most important points (Item 41). It appears that content-oriented examinations have a negative effect on creative growth, and that the encouragement of student comments in large classes may detract from the instructor's ability to employ more potent techniques (e. g., explanations of criticisms, introduction of stimulating ideas).

Table 7 summarizes results obtained for the objective, *Developing skill in expressing myself orally or in writing*. Although this objective is almost always central to introductory courses in speech and expository writing, its more general relevance is demonstrated by the fact that 32.9 percent of those participating in the pilot study selected it as "Important" or "Essential".

TABLE 7

Methods Items Selected as Relevant to "Communication Skills" In Small, Medium, and Large Classes*

Methods Item	Class Size					
	Small		Medium		Large	
	b	r	b	r	b	r
1. Promoted student-teacher discussion	--	--	14	58	--	--
11.Explained criticisms	20	52	24	60	27	61
12.Gave unclear exam questions	-14	-40	--	--	-12	-38
15.Stimulated intellectual effort	22	54	--	--	26	61
17.Explained clearly	--	--	17	52	--	--
40.Tested important points	-13	35	-22	32	-29	36
41.Assessments required originality	54	68	69	77	65	77
N	2658		10,033		1559	
R^2	.546		.656		.668	

* N for "Very Large" classes was 53, too low to permit useful analysis.
Note: Decimals associated with "b" and "r" have been omitted.
Multiple regression equations showed that student reports of instructional methods accounted for 55% to 67% of the variance in progress ratings on this objective. Three methods items were included for all class sizes--Item 11 ("Explained the reasons for criticisms of students" academic performance"), Item 40, negatively ("Gave tests, projects, etc., that covered important points of the course") and Item 41 ("Gave projects, tests, or assignments that required original or creative thinking"). All three items have to do with appraisal and feedback; they suggest that progress on this objective is facilitated when these are performed creatively and conscientiously; instructors whose examinations focus on the most important points are generally less effective in promoting progress on this objective, perhaps because this focus inhibits the kind of creative appraisal which promotes growth.

Attention should be paid to the relative size of the regression (b) weights shown in Table 8. The weight associated with Item 41 (tests, projects, assignments requiring original or creative thinking) is more than twice that associated with any other item. It appears that this skill may be promoted by activities and feedback which promote an effective style of communication more than standardized rules and procedures.

Two items were selected only for medium sized classes [Item 1, "Promoted teacher-student discussion (as opposed to mere responses to questions)"; and Item 17, "Explained course material clearly, and explanations were to the point"]. Two others were included for small and large classes, but not those of medium size (Item 12, negatively--Gave examination questions which were unclear"; and Item 15--"Stimulated students to intellectual effort beyond that required by most courses"). The peculiarities of item redundancies may best explain these findings.

For example, Item 15 (stimulated effort, small and large classes) and Item 1 (promoted discussion, medium size classes) overlap significantly (i. e., correlated .60 with each other); therefore, it would be unlikely that both would be retained in a multiple regression equation. Item 1 was correlated with progress ratings slightly higher in medium sized classes (.58) than it was in small and large classes (.55 and .57, respectively). But for Item 15, the correlations were reversed (.54 and .61 for small and large classes; .53 for medium sized classes). These small statistical differences probably account for the differential pattern of selection. One may speculate that, in any sized class, the stimulation of student effort promotes progress in developing communication skills; frequently, this can be done by promoting student-teacher discussion.

The objective, *Gaining a broader understanding and appreciation of intellectual-cultural activity (music, science, literature, etc.)* is especially appropriate for broad survey classes. Since these have become less popular in institutions of higher education, it has been chosen relatively infrequently and was originally on the list of objectives to be excluded from the revised IDEA form. On the basis of user feedback which argued for its relative uniqueness and cruciality to those teaching survey courses, it was reinstated. The results of multiple regression analysis for this objective are given in Table 8.

Progress ratings on this objective were less accurately predicted than for any other objective; ratings of instructional methods accounted for only 45%, 52%, and 58% of the variance in progress ratings for small, medium, and large-sized classes--significantly less than that reported for the other objectives (Tables 2-7).
This may reflect the omission of important teaching methods on the IDEA form; it may also reflect the difficulty students have in rating their progress on this objective.

Three items were selected across all class sizes--Item 18 (negatively), "Related course material to real life situations"; Item 20, "Introduced stimulating ideas about the subject"; and Item 41, "Gave projects, tests, or assignments that required original or creative thinking". The regression weight, *b*, for Item 20 was much larger than for any other item, underscoring the importance of introducing stimulating ideas. At the same

TABLE 8

Methods Items Selected as Relevant to "Broad Intellectual/Cultural Development" In Small, Medium, and Large Classes*

Methods Item	Class Size					
	Small		Medium		Large	
	b	r	b	r	b	r
2. Helped students answer own questions	--	--	--	--	-19	45
4. Seemed enthusiastic	--	--	26	53	--	--
7. Spoke with expressiveness	13	47	--	--	--	--
11.Explained criticisms	18	47	--	--	--	--
12.Gave unclear examination questions	--	--	--	--	-19	-49
16.Clearly stated course objectives	--	--	--	--	37	51
18.Related course to real life	-43	23	-50	26	-61	17
20.Introduced stimulating ideas	72	59	81	62	97	59
40.Tested important points	--	--	--	--	-31	30
41.Assessments required originality	21	49	29	55	26	49
N	1256		4829		1140	
R^2	.451		.521		.584	

* N for "Very Large" classes was 57, too low to permit useful analysis.
Note: Decimals associated with "b" and "r" have been omitted.

time, progress is likely to be inhibited if emphasis is given to relating course content to everyday life; any "gains" to be had from this technique will be more than offset by "losses" incurred from limiting implementation of more potent educational procedures. Assessing original/creative thinking contributed to progress ratings on this objective, as it did for several other objectives where learning appears to be more "personalized" (applications for thinking, problem solving; creative capacities; personal responsibility; communication skills).

Significant contributions to the prediction of progress ratings on this objective were also made by two additional items for small classes, one for medium sized classes, and four for large classes. Since these seven items were each selected for only one class size , interpretations are difficult and may be misleading.

Using Existing Methods Items to Predict Progress Ratings on "Proposed" Objectives. As noted previously, of the six proposed objectives, five were tentatively selected for inclusion on the revised IDEA form. In this section, step-wise multiple regression procedures are used to explore the utility of existing "methods" items for predicting progress ratings on these five objectives.

Data for these analyses came from the pilot group. The number of classes included was markedly reduced from that used in the previous analyses because (1) there were only 3668 classes in the pilot experiment, (2) only approximately one-half of these contained information regarding a given proposed objective[12], and (3) only those classes for which the instructor selected the objective as "Important" or "Essential" were included. These considerations precluded performing a separate analysis for each class size; therefore, only medium sized classes (the most common size, enrolling 15-34 students) were used.

Table 9 shows the partial regression weights for items which made significant independent contributions to the prediction of each progress rating as well as the percent of variance accounted for by each (beyond that accounted for by items previously selected through the step-wise procedure).

In general, ratings on existing methods items were relatively unsuccessful in predicting progress ratings on the proposed objectives; except for the original objective of *Broad Intellectual/Cultural Development*, R^2 values were consistently lower than they were for existing objectives. For three objectives (*Team Skills; Finding Learning Resources*; and *Critical Analysis*), R^2 was below .40. Diagnostic assistance to improve progress ratings on these objectives will require data about approaches to teaching beyond those depicted on the original IDEA form.

Conclusions Regarding the Utility of "Existing" Methods Items

Tables 2-9 reported partial regression weights for each of the 22 original methods items as predictors of 12 criteria (progress ratings on 7 existing objectives and 5 proposed objectives). For four of the existing objectives, a separate analysis was conducted for small, medium, large, and very large classes; for the other three, only three class sizes were employed because of the limited number of "very large" classes. For the five proposed (new) objectives, only medium sized classes were employed; but three multiple regression analyses were conducted [each using a statistically determined combination of old items and five of the new (proposed) methods items]. Therefore, the statistical process of selecting methods items from the existing list of 22 involved a total of 40 multiple regressions.

In determining whether to retain or discard the 22 existing methods items, it is relevant to know how often each made an independent contribution to the prediction of progress ratings on these

[12]Six sets of tryout items were prepared, each of which included three of the six proposed objectives and five of the proposed fifteen methods. As a consequence, each objective was included in only three of the six sets, and each proposed method was included in only two of the six sets.

12 objectives. Table 10 classifies items in accordance with (a) the number of times their partial regression weights were statistically significant and (b) the number of times their partial regression weight was among the top three in explaining variation in progress ratings.

TABLE 9

Utility of Existing Methods for Predicting Progress Ratings On Five Proposed Objectives (Medium Sized Classes Only)

"Method" Item	Proposed Objective									
	Team Skills		Finding Resources		Seeking Learning		Critical Analysis		Values Develop.	
	b	% Var	b	% Var	b	% Var	b	% Var	b	% Var
1. Promoted discussion					.17	0.8	.35	28.1	.21	3.2
2. Helped Ss answer own Qs			.30	18.7	.31	51.1			.15	0.8
3. Encouraged free expressn									-.13	0.4
4. Seemed enthusiastic					-.11	0.5				
5. Changed approaches	.81	3.4			.10	0.2				
6. Exams required rote	.33	4.0			.06	0.3			.14	0.9
7. Expressive speech							-.17	1.0		
8. Demo importance of subj.	.43	1.0			.17	1.0				
9. Dry and dull lectures										
10.Clear fit of topics									.09	0.2
11.Explained criticism										
12.Unclear exam questions										
13.Encouraged S comments										
14.Summarized effectively										
15.Stimulated intellectl effort					.15	2.4	.17	1.9		
16.Clear objectives										
17.Explained clearly									.08	0.4
18.Related to real life									.13	1.5
19.Picky exams										
20.Stimulating ideas	-.79	3.6					.25	0.5	.15	45.3
40.Tested important points			.56	2.8	.14	0.6			-.10	0.2
41.Exams require creativity	.72	21.1	.15	1.6	.07	0.4	.20	4.8	1.6	1.3
N	317		225		1002		682		954	

R^2	.332	.230	.574	.362	.542

TABLE 10

Frequency With Which Existing Methods Items Made Independent Contributions To the Prediction of Progress Ratings, Based on 40 Multiple Regression Analyses

Frequency of Statistical Significance	**Items**
0-4 Analyses	1, 3, 5, 6, 7, 9, 12, 14, 16, 19
5-9 Analyses	8, 13, 17
10-14 Analyses	2, 4, 10, 11, 40
15 or More Analyses	15, 18, 20, 41
Frequency as a "Top Three" Predictor	
0-1 Times	3, 4, 6, 9, 13, 14, 17, 19
2-5 Times	1, 5, 7, 8, 12, 16
6-10 Times	2, 11
11 or More Times	10, 15, 18, 20, 40, 41

On the basis of the results shown in Table 10, existing items were grouped as follows:

Items with obvious value (retain):

2. "Found ways to help students answer their own questions".
8. "Demonstrated the importance and significance of the subject matter".
10. "Made it clear how each topic fit into the course".
11. "Explained the reasons for criticisms of students' academic performance".
15. "Stimulated students to intellectual effort beyond that required by most course".
18. "Related course material to real life situations".
20. "Introduced stimulating ideas about the subject".
40. "Gave tests, projects, etc. that covered important points of the course".
41. "Gave projects, tests, or assignments that required original or creative thinking".

Items with little value (eliminate):

3. "Encouraged students to express themselves freely and openly".
6. "Gave examinations which stressed unnecessary memorization".
9. "Made presentations which were dry and dull".
14. "Summarized material in a manner which aided retention".
19. "Gave examination questions which were unreasonable detailed (picky)".

Items of some value (reconsider after all analyses are complete):

1. "Promoted teacher-student discussion (as opposed to mere responses to questions)".
4. "Seemed enthusiastic about the subject matter".
5. "Changed approaches to meet new situations".

7. "Spoke with expressiveness and variety in tone of voice".
12. "Gave examination questions which were unclear".
13. "Encouraged student comments even when they turned our to be incorrect or irrelevant".
16. "Clearly stated the objectives of the course".
17. "Explained course material clearly, and explanations were to the point".

Using Proposed Methods Items to Predict Progress on "Existing" Objectives. A total of 15 new methods items were included in the pilot study. Since only five of these was administered by each participating institution, it was possible to collect statistical information about each item in only about one-third of the pilot classes. The number of classes available for analysis was further limited by (a) the decision to study only medium-sized (15-34) classes and (b) the requirement that the instructor chose the objective in question as "Important" or "Essential".

These limitations imposed restrictions on the statistical treatment of the data. Obviously, it was not possible to produce a matrix of inter-correlations among all 15 of the proposed items; the best we could do was to analyze five items at a time.

Further, since our interest was in knowing whether or not the proposed items would add to the information available from existing items, it was desirable to consider both "existing" and "proposed" methods items. However, because of the limited number of classes available for analysis, stable results could not be obtained by a simultaneous analysis of all 27 items (22 "existing" plus 5 "proposed"). To reduce the number of independent variables, predictions of progress ratings for a given objective were made (using the regression equation previously developed for middle-sized classes). This prediction became the first independent variable in a six-variable multiple regression analysis. Mean ratings for the first five tryout methods items constituted the other five independent variables. The dependent variable, of course, was the mean progress rating on the objective in question. Because of the limited number of classes, independent variables whose partial regression weight was significant at the .05, rather than the .01 level, were retained.

Three such analyses were conducted for each of eight existing objectives--one for each of the three sets of pilot items. Results are displayed in Tables 11-17.

Because the number of classes available was relatively small, the stability of the estimated correlation and regression coefficients is in doubt, even though their statistical significance is not. It seems imprudent to make detailed interpretations of each table until the database is large enough to ensure statistical stability.

Still, some broad conclusions can be drawn. First, for these eight "old" objectives, the proposed (new) methods items added only modestly to the accuracy of prediction obtained from the existing 22 items. On average, the existing (old) items accounted for 61.5% of the variance in progress ratings on existing objectives; each of the three sets of five "new" methods items added an average of 4.4% to this figure.

There were some differences among the eight objectives as well as among the three sets of new (proposed) items. For Set 1 (Items 60-64), the average percent increase in variance accounted for was 6.25%; for Sets 2 and 3, this figure was about 3.5%. The proposed items made their biggest contribution to predicting progress ratings on *Developing skill in expressing myself orally and in writing*, where the first two sets of new items each added 12 percent to the variance accounted for by the original 22 methods items.

TABLE 11

Predicting Progress Ratings on "21. Gaining Factual Knowledge. . ."
From Both Existing and Proposed Teaching Methods
Try-Out Sample, Medium Sized Classes

Analysis	Independent Variable	b	r
1	1. Predicted Rating Based on 22 Existing Items	.827	.78
	60. Displayed personal interest		
	61. Formed teams, discussion groups	-.042	.07
	62. Involved students with "hands on"	.115	.30
	63. Used variety of assessment methods		
	64. Scheduled in a helpful manner	.105	.54
	N	561	
	R^2	.671	
2	1. Predicted Rating Based on 22 Existing Items	1.097	.83
	65. Encouraged use of multiple resources	-.080	.13
	66. Asked students to help each other		
	67. Shared personal experiences with class		
	68. Inspired students to set challenging goals		
	69. Frequent, timely feedback		
	N	140	
	R^2	.701	
3	1. Predicted Rating Based on 22 Existing Items	1.074	.82
	70. Encouraged outside student-faculty interaction		
	71. Established high achievement expectations		
	72. Asked students to share experiences with diverse others	-.063	.34
	73. Provided incentives to complete assignments		
	74. Required students be responsible for their learning	.087	.48
	N	352	
	R^2	.698	

TABLE 12

Predicting Progress Ratings on "22. Learning Fundamental Principles. . ."
From Both Existing and Proposed Teaching Methods
Try-Out Sample, Medium Sized Classes

Analysis	Independent Variable	b	r
1	1. Predicted Rating Based on 22 Existing Items	.928	.79
	60. Displayed personal interest	.059	.56
	61. Formed teams, discussion groups	-.047	.09
	62. Involved students with "hands on"	.059	.28
	63. Used variety of assessment methods		
	64. Scheduled in a helpful manner		
	N	652	
	R^2	.655	
2	1. Predicted Rating Based on 22 Existing Items	1.200	.84
	65. Encouraged use of multiple resources		
	66. Asked students to help each other		
	67. Shared personal experiences with class	.080	.49
	68. Inspired students to set challenging goals	-.161	.51
	69. Frequent, timely feedback		
	N	140	
	R^2	.720	
3	1. Predicted Rating Based on 22 Existing Items	1.030	.82
	70. Encouraged outside student-faculty interaction		
	71. Established high achievement expectations	.078	.48
	72. Asked students to share experiences with diverse others		
	73. Provided incentives to complete assignments		
	74. Required students be responsible for their learning		
	N	301	
	R^2	.698	

TABLE 13

Predicting Progress Ratings on "23. Applications for Thinking, Problem Solving. . ."
From Both Existing and Proposed Teaching Methods
Try-Out Sample, Medium Sized Classes

Analysis	Independent Variable	b	r
1	1. Predicted Rating Based on 22 Existing Items	.770	.78
	60. Displayed personal interest		
	61. Formed teams, discussion groups		
	62. Involved students with "hands on"	.034	.41
	63. Used variety of assessment methods		
	64. Scheduled in a helpful manner	.133	.57
	N	597	
	R^2	.632	
2	1. Predicted Rating Based on 22 Existing Items	.878	.85
	65. Encouraged use of multiple resources		
	66. Asked students to help each other		
	67. Shared personal experiences with class		
	68. Inspired students to set challenging goals		
	69. Frequent, timely feedback	.151	.56
	N	151	
	R^2	.729	
3	1. Predicted Rating Based on 22 Existing Items	1.083	.84
	70. Encouraged outside student-faculty interaction		
	71. Established high achievement expectations		
	72. Asked students to share experiences with diverse others		
	73. Provided incentives to complete assignments		
	74. Required students be responsible for their learning		
	N	250	
	R^2	.710	

TABLE 14

Predicting Progress Ratings on "24. Professional Skills, Viewpoints"
From Both Existing and Proposed Teaching Methods
Try-Out Sample, Medium Sized Classes

Analysis	Independent Variable	b	r
1	1. Predicted Rating Based on 22 Existing Items	.689	.70
	60. Displayed personal interest		
	61. Formed teams, discussion groups	-.074	.19
	62. Involved students with "hands on"	.189	.45
	63. Used variety of assessment methods		
	64. Scheduled in a helpful manner	.144	.57
	N	457	
	R^2	.593	
2	1. Predicted Rating Based on 22 Existing Items	.690	.82
	65. Encouraged use of multiple resources		
	66. Asked students to help each other	.137	.49
	67. Shared personal experiences with class	.148	.67
	68. Inspired students to set challenging goals		
	69. Frequent, timely feedback		
	N	82	
	R^2	.706	
3	1. Predicted Rating Based on 22 Existing Items	1.082	.84
	70. Encouraged outside student-faculty interaction		
	71. Established high achievement expectations		
	72. Asked students to share experiences with diverse others	-.097	.43
	73. Provided incentives to complete assignments		
	74. Required students be responsible for their learning	.142	.49
	N	275	
	R^2	.760	

TABLE 15

Predicting Progress Ratings on "26. Creative Capacities"
From Both Existing and Proposed Teaching Methods
Try-Out Sample, Medium Sized Classes

Analysis	Independent Variable	b	r
1	1. Predicted Rating Based on 22 Existing Items	.670	.85
	60. Displayed personal interest	.120	.62
	61. Formed teams, discussion groups	-.050	.46
	62. Involved students with "hands on"	.196	.61
	63. Used variety of assessment methods		
	64. Scheduled in a helpful manner		
	N	158	
	R^2	.774	
2	1. Predicted Rating Based on 22 Existing Items	1.099	.88
	65. Encouraged use of multiple resources		
	66. Asked students to help each other		
	67. Shared personal experiences with class		
	68. Inspired students to set challenging goals		
	69. Frequent, timely feedback		
	N	56	
	R^2	.768	
3	1. Predicted Rating Based on 22 Existing Items	.991	.87
	70. Encouraged outside student-faculty interaction	.147	.33
	71. Established high achievement expectations		
	72. Asked students to share experiences with diverse others		
	73. Provided incentives to complete assignments		
	74. Required students be responsible for their learning		
	N	76	
	R^2	.793	

TABLE 16

Predicting Progress Ratings on "28. Broad Intellectual/Cultural Development"
From Both Existing and Proposed Teaching Methods
Try-Out Sample, Medium Sized Classes

Analysis	Independent Variable	b	r
1	1. Predicted Rating Based on 22 Existing Items	.842	.71
	60. Displayed personal interest		
	61. Formed teams, discussion groups		
	62. Involved students with "hands on"		
	63. Used variety of assessment methods	.118	.46
	64. Scheduled in a helpful manner		
	N	128	
	R^2	.540	
2	1. Predicted Rating Based on 22 Existing Items	1.180	.78
	65. Encouraged use of multiple resources		
	66. Asked students to help each other		
	67. Shared personal experiences with class		
	68. Inspired students to set challenging goals		
	69. Frequent, timely feedback		
	N	49	
	R^2	.680	
3	1. Predicted Rating Based on 22 Existing Items	1.272	.69
	70. Encouraged outside student-faculty interaction	-.210	.24
	71. Established high achievement expectations	.483	.34
	72. Asked students to share experiences with diverse others	-.540	.46
	73. Provided incentives to complete assignments		
	74. Required students be responsible for their learning		
	N	80	
	R^2	.547	

TABLE 17

Predicting Progress Ratings on "29. Communication Skills"
From Both Existing and Proposed Teaching Methods
Try-Out Sample, Medium Sized Classes

Analysis	Independent Variable	b	r
1	1. Predicted Rating Based on 22 Existing Items	.585	.72
	60. Displayed personal interest	.165	.48
	61. Formed teams, discussion groups	.212	.59
	62. Involved students with "hands on"		
	63. Used variety of assessment methods		
	64. Scheduled in a helpful manner	.125	.44
	N	367	
	R^2	.638	
2	1. Predicted Rating Based on 22 Existing Items	1.129	.83
	65. Encouraged use of multiple resources	.179	.59
	66. Asked students to help each other		
	67. Shared personal experiences with class		
	68. Inspired students to set challenging goals	-.172	.54
	69. Frequent, timely feedback	-.097	.34
	N	115	
	R^2	.813	
3	1. Predicted Rating Based on 22 Existing Items	.829	.79
	70. Encouraged outside student-faculty interaction		
	71. Established high achievement expectations	.278	.29
	72. Asked students to share experiences with diverse others		
	73. Provided incentives to complete assignments		
	74. Required students be responsible for their learning		
	N	146	
	R^2	.638	

One way to judge the value of a proposed "methods" items is the number of times it added to the variance in progress ratings "explained" by existing methods items. By this criterion, nine of these items made independent contributions in predicting at least three of the seven progress ratings:

60. "Displayed a personal interest in me and my learning○--significant in 3 analyses.
61. "Formed "teams" or "discussion groups" to facilitate learning"--significant in 5 analyses.
62. "Involved students in "hands on" projects such as research, case studies, or real life"--significant in 5 analyses.
64. "Scheduled course work (class activities, tests, projects) in ways which encouraged students to stay up to date in their work"--significant in 4 analyses.
65. "Encouraged students to use multiple resources (e. g., data banks, library holdings, outside expects) to improve understanding"--significant in 3 analyses.
68. "Inspired students to set and achieve goals which really challenged them"--significant in 3 analyses.
71. "Established high expectations of achievement in this class"--significant in 4 analyses.
72. "Asked students to share ideas and experiences with others whose backgrounds and viewpoints differ from their own"--significant in 3 analyses.
74. "Required students to take their share of responsibility for learning"--significant in 3 analyses.

Three other items made significant additions to the prediction of two progress ratings:

67. "Shared relevant personal experiences, attitudes, and values with class".
69. "Provided timely and frequent feedback on tests, reports, projects, etc. to help students improve".
70. "Encouraged student-faculty interaction outside of class (office visits, phone calls, e-mail, etc"

Two items were significant in only one analysis:

63. "Relied on a variety of methods, not only tests, to evaluate student progress on course objectives."
66. "Asked students to help each other understand ideas or concepts".

And one item was not significant in any of the eight analyses:

73. "Created incentives that facilitated the timely completion of course assignments".

Of course, the value of the proposed "methods" items also should be judged by their success in predicting the "new" (proposed) objectives. This question is explored in the next section.

Using Proposed Methods Items to Predict Progress on Proposed Objectives. There were 5 new objectives which had been tentatively selected to be included on the revised form. Again, the predictive potency of the proposed (new) method items could be examined only in three sets of five items each, meaning that information about each item was available in only about one-third of the tryout classes. Again, the number of available classes was further limited by (a) the decision to study only medium-sized (15-34) classes and (b) the requirement that the instructor chose the objective in question as "Important" or "Essential".

Again, our interest was in knowing whether or not the proposed items would add to the information provided by existing items. The procedure followed in the previous section was employed. Progress ratings for a

given objective were first predicted on the basis of responses to existing methods items, using the previously developed regression equation for middle-sized classes. This prediction became the first independent variable in a six-variable multiple regression analysis. Mean ratings for the five tryout methods items constituted the other five independent variables. The dependent variable, of course, was the mean progress rating on the objective in question. Three analyses were performed for each proposed objective, one for each of the three sets of "new" methods items. Results are shown in Tables 18-22.

As shown in Table 18, the "old" methods items were relatively unsuccessful in accounting for variation in progress ratings on *Acquiring Team Skills;* however, two of the proposed ("new") methods were substantially related to this criterion: 61. "Formed" teams' or "discussion groups' to facilitate learning"; and 65. "Encouraged students to use multiple resources (e. g., data banks, library holdings, outside experts) to improve understanding". Five of the other "new" items also made a significant contribution to the prediction of progress ratings on this objective (60, 64, 68, 72, and 74), but the regression coefficient for two of these (60, 68) was negative.

Progress ratings on proposed Objective 51 (*Learning how to find and use resources for answering questions or solving problems*) were significantly related to the best combination of ratings on the 22 existing methods; see Table 19. But the zero order correlations of several of the "new" methods with these progress ratings were at least as high as the R for old items--62. "Involved students in "hands on" projects such as research, case studies, or "real life" activities'; 64. "Scheduled course work (class activities, tests, projects) in ways which encouraged students to stay up to date in their work"; 65. "Encouraged students to use multiple resources (e. g., data banks, library holdings, outside experts) to improve understanding"; 68. "Inspired students to set and achieve goals which really challenged them"; and 71. "Established high expectations of achievement in this class". Item 69 also made an independent contribution to the prediction of these progress ratings.

For the third proposed objective (*53. Acquiring an interest in learning more by asking questions and seeking answers*), the best combination of old methods items was consistently the best predictor of progress ratings. However, three of the proposed "new" methods made substantial independent contributions (60. "Displayed a personal interest in me and my learning"; 66. "Asked students to help each other understand ideas or concepts"; and, negatively, 68. "Inspired students to set and achieve goals which really challenged them". See Table 20. The negative value associated with Item 68 does not mean that ratings on this item correlated negatively with progress ratings; in fact, the two were positively correlated (.63). Its negative regression weight probably means that instructors who expend time and effort encouraging students to set challenging goals may have to forgo the opportunity to employ more potent techniques (e. g., getting students to help each other; encouraging them to answer their own questions). Several other proposed items made small independent contributions to the prediction of progress ratings on this new objective, including items 63, 64, 67, 70, 71, and 73.

Table 21 explores the utility of "old" and "new" methods items in predicting progress ratings on *Learning to analyze and critically evaluate ideas, arguments, and points of view.* The existing methods items were relatively successful in predicting these criterion ratings. But two of the proposed methods also made substantial independent contributions: 70. "Encouraged student-faculty interaction outside of class (office visits, phone calls, e-mail, etc.)" and 74. "Required students to take their share of responsibility for learning". Small, but significant, contributions were also made by "new" items 61, 62, 64, 66, 67, 69, and 72. For items 62 and 67, the regression coefficient was negative even though the zero-order correlations were positive.

TABLE 18

Predicting Progress Ratings on *Acquiring Team Skills*
From Both Existing and Proposed Teaching Methods
Try-Out Sample, Medium Sized Classes

Analysis	Independent Variable	b	r
1	1. Predicted Rating Based on 22 Existing Items		
	60. Displayed personal interest	-.36	.32
	61. Formed teams, discussion groups	.66	.86
	62. Involved students with "hands on"		
	63. Used variety of assessment methods		
	64. Scheduled in a helpful manner	.31	.47
	N	48	
	R^2	.743	
2	1. Predicted Rating Based on 22 Existing Items		
	65. Encouraged use of multiple resources	1.00	.76
	66. Asked students to help each other		
	67. Shared personal experiences with class		
	68. Inspired students to set challenging goals	-.26	.19
	69. Frequent, timely feedback		
	N	48	
	R^2	.608	
3	1. Predicted Rating Based on 22 Existing Items	.73	.63
	70. Encouraged outside student-faculty interaction		
	71. Established high achievement expectations		
	72. Asked students to share experiences with diverse others	.13	.53
	73. Provided incentives to complete assignments		
	74. Required students be responsible for their learning	.36	.47
	N	211	
	R^2	.459	

TABLE 19

Predicting Progress Ratings on *Finding, Using Resources to Answer Own Questions*
From Both Existing and Proposed Teaching Methods
Try-Out Sample, Medium Sized Classes

Analysis	Independent Variable	b	r
1	1. Predicted Rating Based on 22 Existing Items	.440	.69
	60. Displayed personal interest		
	61. Formed teams, discussion groups		
	62. Involved students with "hands on"	.259	.69
	63. Used variety of assessment methods		
	64. Scheduled in a helpful manner	.281	.68
	N	41	
	R^2	.616	
2	1. Predicted Rating Based on 22 Existing Items	.440	.65
	65. Encouraged use of multiple resources	.444	.71
	66. Asked students to help each other		
	67. Shared personal experiences with class	-.274	.39
	68. Inspired students to set challenging goals	.210	.69
	69. Frequent, timely feedback	.124	.49
	N	66	
	R^2	.722	
3	1. Predicted Rating Based on 22 Existing Items	.613	.66
	70. Encouraged outside student-faculty interaction		
	71. Established high achievement expectations	.462	.59
	72. Asked students to share experiences with diverse others		
	73. Provided incentives to complete assignments		
	74. Required students be responsible for their learning		
	N	102	
	R^2	.501	

TABLE 20

Predicting Progress Ratings on *Increase Interest in Learning*
From Both Existing and Proposed Teaching Methods
Try-Out Sample, Medium Sized Classes

Analysis	Independent Variable	b	r
1	1. Predicted Rating Based on 22 Existing Items	.643	.80
	60. Displayed personal interest	.247	.74
	61. Formed teams, discussion groups		
	62. Involved students with "hands on"		
	63. Used variety of assessment methods	.047	.58
	64. Scheduled in a helpful manner	.057	.65
	N	614	
	R^2	.686	
2	1. Predicted Rating Based on 22 Existing Items	.889	.83
	65. Encouraged use of multiple resources		
	66. Asked students to help each other	.207	.69
	67. Shared personal experiences with class	.066	.60
	68. Inspired students to set challenging goals	-.213	.63
	69. Frequent, timely feedback		
	N	135	
	R^2	.731	
3	1. Predicted Rating Based on 22 Existing Items	1.126	.88
	70. Encouraged outside student-faculty interaction	.142	.72
	71. Established high achievement expectations	-.098	.63
	72. Asked students to share experiences with diverse others		
	73. Provided incentives to complete assignments	-.083	.58
	74. Required students be responsible for their learning		
	N	245	
	R^2	.786	

TABLE 21

Predicting Progress Ratings on *Critical Analysis*
From Both Existing and Proposed Teaching Methods
Try-Out Sample, Medium Sized Classes

Analysis	Independent Variable	b	r
1	1. Predicted Rating Based on 22 Existing Items	.877	.77
	60. Displayed personal interest		
	61. Formed teams, discussion groups	.081	.46
	62. Involved students with "hands on"	-.068	.44
	63. Used variety of assessment methods		
	64. Scheduled in a helpful manner	.135	.54
	N	437	
	R^2	.625	
2	1. Predicted Rating Based on 22 Existing Items	.915	.83
	65. Encouraged use of multiple resources		
	66. Asked students to help each other	.182	.69
	67. Shared personal experiences with class	-.127	.57
	68. Inspired students to set challenging goals		
	69. Frequent, timely feedback	.178	.62
	N	97	
	R^2	.731	
3	1. Predicted Rating Based on 22 Existing Items	.518	.71
	70. Encouraged outside student-faculty interaction	.225	.77
	71. Established high achievement expectations		
	72. Asked students to share experiences with diverse others	.138	.61
	73. Provided incentives to complete assignments		
	74. Required students be responsible for their learning	.252	.59
	N	104	
	R^2	.633	

The final proposed objective, *Developing a clearer understanding of, and commitment to, personal values* is drawn from the affective, domain. Results are shown in Table 22. Again, the existing methods items were

TABLE 22

Predicting Progress Ratings on *Values Development*
From Both Existing and Proposed Teaching Methods
Try-Out Sample, Medium Sized Classes

Analysis	Independent Variable	b	r
1	1. Predicted Rating Based on 22 Existing Items	.781	.74
	60. Displayed personal interest	.151	.61
	61. Formed teams, discussion groups	.044	.45
	62. Involved students with "hands on"		
	63. Used variety of assessment methods		
	64. Scheduled in a helpful manner		
	N	677	
	R^2	.568	
2	1. Predicted Rating Based on 22 Existing Items	.791	.72
	65. Encouraged use of multiple resources		
	66. Asked students to help each other		
	67. Shared personal experiences with class	.277	.64
	68. Inspired students to set challenging goals		
	69. Frequent, timely feedback	-.149	.38
	N	70	
	R^2	.571	
3	1. Predicted Rating Based on 22 Existing Items	.571	.77
	70. Encouraged outside student-faculty interaction	.191	.61
	71. Established high achievement expectations	-.150	.60
	72. Asked students to share experiences with diverse others	.206	.75
	73. Provided incentives to complete assignments		
	74. Required students be responsible for their learning	.218	.60
	N	196	
	R^2	.695	

significantly related to progress ratings on this objective. But four of the proposed new methods made substantial independent contributions to their prediction: 67. "Shared personal experiences, attitudes, and values with the class"; 70. "Encouraged student-faculty interaction outside of class (office visits, phone calls, e-mail, etc.)"; 72. "Asked students to share ideas and experiences with others whose backgrounds and viewpoints differ from their own"; and 74. "Required students to take their share of responsibility for learning". Four other items made smaller independent contributions (items 60, 61, 69, and 71). Partial regression coefficients for the latter two of these were negative even though their zero-order correlations with progress ratings on this objective were positive (.38 and .60, respectively).

Conclusions Regarding the Utility of "Proposed" Methods Items

The value of each of the 15 proposed (new) methods items for making an independent contribution to the prediction of progress ratings is shown in Tables 12-19 (for the seven existing objectives) and Tables 20-24 (for the five proposed objectives). The first step in deciding whether to retain or discard an item was to examine the frequency with which it was significant in these 12 multiple regression analyses. It is also of interest to know how often it was one of the top two predictors.[13] Table 23 summarizes these frequency tallies.

TABLE 23

Frequency With Which Proposed Methods Items Made Independent Contributions To the Prediction of Progress Ratings, Based on 12 Multiple Regression Analyses

Frequency of Statistical Significance	**Items**
0 or 1 Analyses	73
2-4 Analyses	63, 65, 66
5-7 Analyses	60, 67, 68, 69, 70, 71, 72, 74
8 or More Analyses	61, 62, 64
Frequency as a "Top Two" Predictor	
None	73
1 Time	63, 69, 72

[13]When existing methods items were examined (Table 11), the frequency with which an item was a "top three" predictor was recorded. In this section, only the "top two" predictors are considered. This is because the proposed items were considered five at a time, rather than all at once as was done for existing methods.

2 or 3 Times	60, 61, 66, 67, 68, 70, 71
4 Times or More	62, 64, 65, 74

On the basis of results reported in Table 23, these items were grouped as follows:

Items with the most obvious value (tentatively retain)

61. "Formed "teams" or "discussion groups" to facilitate learning".
62. "Involved students in "hands on" projects such as research, case studies, or "real life" activities"
64. "Scheduled course work (class activities, tests, projects) in ways which encouraged students to stay up to date".
68. "Inspired students to set and achieve goals which really challenged them".
70. "Encouraged student-faculty interaction outside of class (office visits, phone calls, e-mail, etc)"

Items with the least obvious value (tentatively reject)

63. "Relied on a variety of methods, not only tests, to evaluate student progress on course objectives".
73. "Created incentives that facilitated the timely completion of course assignments".

Items with intermediate value (reconsider using additional criteria)

60. "Displayed a personal interest in me and my learning".
65. "Encouraged students to use multiple resources (e.g., data banks, library holdings, outside experts) to improve understanding".
66. "Asked students to help each other understand ideas or concepts".
67. "Shared personal experiences, attitudes, and values with the class".
69. "Provided timely and frequent feedback on tests, reports, projects, etc. to help students improve".
71. "Established high expectations of achievement for this class".
72. "Asked students to share ideas and experiences with others whose backgrounds and viewpoints differ from their own".
74. "Required students to take their share of responsibility for learning".

Synthesizing the Selection of Methods Items. In a previous section, tentative conclusions were drawn about the utility of existing items for predicting progress ratings on both existing (old) objectives and proposed (new) objectives; see *Conclusions* from Table 11. In the section just completed similar conclusions were drawn about the utility of proposed (new) methods items. Since it was predetermined that the revised IDEA form should contain the same number of methods items as the old form (20), the problem became one of synthesizing these two sets of conclusions.

Two additional considerations were regarded as relevant.

(1) Major dimensions of teaching effectiveness be adequately represented.

(2) To enhance instructional improvement efforts, as many "relevant" methods (those related to progress ratings on each objective) as possible should be identified.

1. Dimensions of Effective Instruction. Each of the 37 items under consideration was classified according to an "Instructional Dimensions" scheme based largely on the Chickering-Gamson *Seven Principles*. The classification system employed five categories:

Student-Faculty Contact, Principle 1 ("Encourage student-faculty contact") in the Chickering-Gamson scheme.

Involving Students, a combination of Chickering-Gamson principles 2 ("Cooperation among

students") and 7 ("Respect diverse talents and ways of learning").

Establishing Expectations, a combination of Chickering-Gamson principles 3 ("Encourage active learning"), 5 ("Emphasize time on task"), and 6 ("Communicate high expectations").

Clarity of Communication, a factor not directly addressed by Chickering-Gamson, but frequently cited in the literature as a key to effective instruction.

Assessment/Feedback, a combination of "assessment" and Chickering-Gamson principle 4 ("Give prompt feedback").

The following chart classifies each of the 37 items by both "Instructional Dimension" and its level of utility as judged by statistical analysis:

	Tentatively Retain	Undecided	Tentatively Reject
Student-Faculty Contact	2, 70 1, 60		
Involving Students	61, 62	4, 65, 66, 72	3, 13
Establishing Expectations	8, 15, 20, 64, 68	7, 71, 74	9, 73
Clarity of Communication	10, 18	5, 16, 17, 67	14
Assessment/Feedback	11, 40, 41	69	6, 12, 19, 63

A total of 14 items are on the "Tentatively Retain" list derived from statistical analyses; 9 others are on the "Tentatively Reject" list. If the latter are excluded, the problem becomes one of choosing 6 items from the remaining 14.

It seemed probable that, if a meaningful assessment was to be made of each of the proposed dimensions, each scale would need to contain at least three items. The chart above shows that this requirement was met on two dimensions--Establishing Expectations (5 items) and Assessment/Feedback (3 items). For this reason, the three "Establishing Expectations" items which were in the "Undecided" column were excluded, leaving 11 from which the remaining 6 must be selected. Choices had to be made from the following:

For *Student-Faculty Contact*: Item 1 and/or Item 60 (Goal: 1 or 2 items).
For *Involving Students*: Item 4, Item 65, Item 66, and Item 72 (Goal: 1-3 items)
For *Clarity of Communication*: Item 5, Item 16, Item 17, Item 67 (Goal: 1-3 items)
For *Assessment Feedback*: Item 69 (Goal: 0-1 item)

2. Comprehensiveness of Diagnostic Information. Lists were prepared of the tentatively retained methods items which were predictive of progress in medium sized classes on each of the 12 objectives chosen for inclusion on the revised IDEA form. The predictive value of the 11 "Undecided" items was also reviewed. The chart below summarizes the findings.

Objective	Tentatively Chosen Items	"Undecided" Items
21. *Factual knowledge. . .*	10, 15, 20, 40, 41, 61, 62, 64	65, 72
22. *Principles, theories. . .*	10, 15, 40, 61, 62, 68	60, 67
23. *Applications. . .*	2, 8, 15, 18, 40, 62, 64	4, 69
24. *Professional skills. . .*	8, 10, 11, 15, 18, 40, 61, 62, 64	66, 67, 72
26. *Creative capacities*	10, 11, 18, 20, 41, 61, 62, 70	5, 60
28. *General/liberal education*	11, 40, 41, 70	1, 16, 17, 72
29. *Communication skills*	11, 40, 41, 61, 64, 68	1, 17, 60, 65, 69

Objective	Tentatively Chosen Items	"Undecided" Items

50. *Team skills*	8, 20, 41, 61, 64, 68	5, 60, 65, 70
51. *Finding resources*	2, 40, 41, 62, 64, 68	65, 67, 69
53. *Increase learning interest*	2, 8, 15, 40, 41, 64, 68, 70	1, 4, 5, 60, 66, 67
54. *Critical analysis*	15, 20, 41, 61, 62, 64, 70	1, 66, 69, 72
55. *Values development*	2, 10, 18, 20, 40, 41, 61, 70	1, 17, 60, 67, 69, 72

To select six items from the "Undecided" list, two standards were applied: (1) preference should be given to those which contributed to the largest number of progress ratings; and (2) preference should be given to those which contributed to the objectives for which only a small number of tentatively chosen items made independent contributions to the prediction of progress ratings. The latter included:

28. *General/liberal education* (4 items)
22. *Principles, theories* (6 items)
29. *Communication skills* (6 items)
50. *Team skills* (6 items)
51. *Finding resources* (6 items)

Items were compared within the instructional dimension to which they belonged, with the following results:

Dimension 1: Student-Faculty Contact	Total Inclusions	"Critical" Inclusions
Item 1	5	2
Item 60	6	3
Dimension 2: Involving Students		
Item 4	2	0
Item 65	4	3
Item 66	3	1
Item 72	5	1
Dimension 4: Clarity of Communication		
Item 5	3	1
Item 16	1	1
Item 17	3	2
Item 67	3	1
Dimension 5: Assessment/Feedback		
Item 69	5	2

Clearly, close choices were involved. Staff consensus was that the following six items be selected:

17. "Explained course material clearly, and explanations were to the point."
60. "Displayed a personal interest in me and my learning."
65. "Encouraged students to use multiple resources (e. g., data banks, library holdings, outside experts) to improve understanding."
66. "Asked students to help each other understand ideas or concepts."
69. "Provided timely and frequent feedback on tests, reports, projects, etc. to help students improve."
72. "Asked students to share ideas and experiences with others whose backgrounds and

viewpoints differ from their own."

These were added to the 14 methods items previously selected on the basis of statistical analysis:

2. "Found ways to help students answer their own questions".
8. "Demonstrated the importance and significance of the subject matter".
10. "Made it clear how each topic fit into the course."
11. "Explained the reasons for criticisms of students' academic performance."
15. "Stimulated students to intellectual effort beyond that required by most courses."
18. "Related course material to real life situations."
20. "Introduced stimulating ideas about the subject."
40. "Gave tests, projects, etc., that covered important points of the course."
41. "Gave projects, tests, or assignments that required original or creative thinking."
61. "Formed "teams" or "discussion groups" to facilitate learning."
62. "Involved students in "hands on" projects such as research, case studies, or "real life" activities."
64. "Scheduled course work (class activities, tests, projects) in ways which encouraged students to stay up to date in their work."
68. "Inspired students to set and achieve goals which really challenged them."
70. "Encouraged student-faculty interaction outside of class (office visits, phone calls, e-mail, etc)"

The revised list of 20 methods and 12 objectives was now complete. In formatting the revised form, they were renumbered. The Appendix relates item numbers on the "original" and "revised" forms.

V. Accounting for Extraneous Influences: Adjusting Scores

Student ratings are influenced by a large number of factors, most of which are reviewed in summary articles by Feldman (1976, 1978, 1979, 1983, 1984) as well as in recent books on evaluating teaching (Braskamp and Ory, 1994; Centra, 1993). A number of these factors represent "biases" in student ratings, and help explain why the validity of such ratings is never perfect. Some "biases" can be directly controlled by the teacher (e. g., being present when ratings are made; providing non-standardized instructions for completing the form; "explaining" the consequences of a high or low rating; exuding enthusiasm; etc.).

But others are not under the instructor's control. It is this set of factors which the IDEA System aspires to take into account. If there are extraneous influences, beyond the instructor's control, which influence student ratings, it is desirable to "level the playing field" so that no one will be advantaged or disadvantaged by such factors.

Research results suggest that the two extraneous circumstances accounted for by the original IDEA System were both relevant. Student ratings were generally more favorable in small, rather than large, classes. And the ratings of highly motivated students were much more favorable than those of unmotivated students. The IDEA System took these findings into account by comparing the ratings for each class with those for similar classes. "Similar" was defined as:

1) in the same size range (Small=under 15; Medium=15-34; Large=35-99; or Very Large (over 99)
(2) in the same "student motivation" group, where motivation was defined by the average response to the question, "I had a strong desire to take this course" (I=less than 3.0; I I= 3.0-3.4; III= 3.5-3.9; IV = 4.0-4.4; and V = 4.5 or higher).

There were three ways in which this method of controlling for extraneous influences was inadequate.

(1) It was insensitive to differences *within* a group; the same comparison group was used for classes with an enrollment of 35 and with an enrollment of 99, or for those with an average motivation rating of 3.51 and 3.99.

(2) It assumed that the rating on "I had a strong desire to take this course" was not under the instructor's control; but the desire to take a case may be related to the instructor's skill and/or reputation, factors which are, at least in part, under her/his control.

(3) It ignored other factors which may be related to student ratings of outcomes; of special concern were (a) the academic work habits of enrollees [of interest because the average rating on "Effort" ("I worked harder on this course than on most courses I have taken") correlated .44 with *Progress on relevant objectives*] and (b) course difficulty which, although correlated only .08 with *Progress on relevant objectives*, was correlated .65 with the "effort" item (Sixbury and Cashin, 1995).

The statistical technique of *multiple regression* made it possible to address the first and third of these shortcomings. The process requires a dependable way of measuring each extraneous variable. Then a weight is calculated for each measure which reflects the degree to which it made an independent contribution to the prediction of a given outcome. The resulting equation can be used to estimate an "expected" outcome given these extraneous circumstances. If this prediction is lower than the obtained results, the instructor was apparently more "successful" than others who contended with similar factors; accordingly, the outcome rating should be "adjusted" upward. Similarly, if the outcome predicted by measures of extraneous circumstances was higher than the obtained results, a downward adjustment should be made. To ensure comparability of numbers, "Adjustments" should be made in ways which guarantee that averages and standard deviations are equal to those obtained for "raw" (unadjusted) ratings[14].

Extraneous measures. The list of potential extraneous circumstances is a long one. For the revision of the diagnostic (long) form, five of these were investigated. Only the first three of these were available to adjust ratings on the short form.

1. *Class size (N)*. Previous research has shown that student ratings are influenced by enrollment. The measure employed was simply the number of students enrolled.

2. *Student course oriented motivation (CM)*. The original IDEA form included experimental item 43, "I really wanted to take this course regardless of who taught it". The average response to this item was used as a measure of course motivation.

3. *Other motivation (M)*. As noted above, the item used to assess motivation in the original IDEA system (Item 36, "I had a strong desire to take this course") is not a pure measure of student motivation from extraneous sources, since instructor reputation and image (factors at least partially under the control of the instructor) may influence the response. To take this into consideration, the average response to Item 42, "I

[14]The process for doing this is as follows: (1) Subtract the predicted outcome average from the obtained outcome average. (2) Multiply this residual by (1 + % of variance accounted for by the measures of extraneous circumstances). (3) Add this result to the grand mean for the outcome.

really wanted to take a course from this instructor", was used to predict the average response to Item 36. Such a prediction describes the portion of the "strong desire to take this course" which was attributable to instructor reputation. By subtracting this prediction from the average response to Item 36, it was possible to obtain a measure of "strong desire to take this course" from which the effect of instructor reputation had been expunged. This residual presumably reflects motivational influences such as time of day the class was offered, the intention of friends to enroll, or the recommendation of a parent or advisor about the desirability of taking the course. The required formula is:

$$M = X_{36} - (1.804711 + .519087\ X_{38})^{15}$$

4. *Discipline-related difficulty (D_2).* Research has suggested that student ratings may be influenced by the difficulty of the course. Difficulty is, in part, inherent in a discipline or in a specific type of course, and therefore is an extraneous influence. For example, for many students, quantitatively-oriented courses in education or the social sciences are frequently perceived as more difficult than those which rely primarily on reading skills. But difficulty is also due, in part, to the way the course is taught--
i. e., to factors which <u>are</u> under the instructor's control. To assess the portion of the difficulty rating which could be assigned to extraneous circumstances, it was necessary first to determine the portion which could be assigned to instructor behavior. Multiple regression was used to predict the average difficulty rating on the basis of teaching methods chosen for retention on the revised form and two other teacher-controlled items [amount of reading; Item 33 (31on the original form); amount of work in other (non-reading) assignments, Item 34 (32 on the original form)]. This prediction (D_1) represented the portion of the difficulty rating which could be traced to factors under the teacher's control. By subtracting it from the average rating of difficulty, a residual was formed which represents that portion of the difficulty rating not under the instructor's control. It was called *Discipline-related difficulty (D_2).* Using item numbers for the revised IDEA form, the formulas needed to develop the D_2 measure are:

$$D_1 = 1.454839 + .386738\ X_{34} + .277309\ X_{33} - .292284\ X_{19} + .137319\ X_2 - .289368\ X_{10}$$
$$D_2 = X_{35} - D_1$$

5. *Student academic effort (E_2).* Students differ significantly in their work habits, self-discipline, and willingness to expend effort on their academic work. Item 35 on the original IDEA form ("I worked harder on this course than on most courses I have taken") was intended to reflect this. As with "difficulty", student effort can be conceptualized as have two components--a part attributable to the instructor, and a second part which is independent of instructor influences. To assess this second component, the same procedure used in assessing "Disciplinary difficulty" was followed. First, the average response to Item 35 was predicted using a multiple regression equation involving the teaching methods and course management items which were retained from the original IDEA form. This was called *Instructor-induced effort (E_1)*. By subtracting this prediction from the obtained average for Item 35, a measure of *Student academic effort (E_2)* which was independent of the instructor's influence was obtained. The required formulas, using item numbers for the revised form, are:

$$E_1 = .096528 + .552277\ X_{34} + .212321\ X_{33} + .362738\ X_4 - .139236\ X_{11}$$
$$E_2 = X_{37} - E_1$$

<u>Adjusting outcome measures</u>. To determine if these measures of potential extraneous circumstances were

[15] All "X" terms refer to the <u>average</u> response. Subscripts identify item numbers on the <u>revised</u> form.

related to outcome measures and, if so, how they should be combined in order to account for the maximum variance in each outcome measure, step-wise multiple regression procedures were again employed. A total of 15 analyses were performed--one for each of the 12 progress ratings on individual objectives, and one for each of three global measures:

(1). The average rating on , *As a result of taking this course, I have more positive feelings toward this field of study.*
(2) *Overall, I rate this instructor an excellent teacher.*
(3) *Overall, I rate this course as excellent.*

For the global measures, ratings were used from all classes in the historical database for which at least 10 ratings were available. For the 12 progress ratings, for both the historical and pilot databases, classes were included only if the instructor had rated the objective as "Important" or "Essential".

Table 24 summarizes the findings for the diagnostic (long) form.

For all of the 15 measures, the partial regression weight for "Course Motivation" (CM) was positive and highly significant. Those teaching classes whose enrollees really wanted to take the course, regardless of who taught it, have a considerable advantage over those with less well motivated students.

The partial regression weight for "Class Size" was significant in eight of the analyses. In each instance, the weight was negative, confirming previous research that, in general, students in small classes report more progress than do those in large classes. Enrollment size was not significantly predictive of the three global ratings; its importance was apparent only in ratings of outcomes on specific objectives.

The "Other Motivation" (M) measure was included in adjustment equations for 9 of the 15 outcome measures. In eight instances, its partial regression weight was negative. This measure was designed to assess motivations other than those associated with course content or the reputation of the instructor, and may include such frivolous matters as when and where the course was taught or whether or not friends were enrolling. Negative partial regression weights would be expected in such cases.

The measure of *Disciplinary difficulty* (D_2) was significantly related to 9 of the 12 progress ratings on specific objectives; for eight of these, its influence was positive. Apparently, more positive outcomes are typically found on courses which are inherently difficult. This finding appears to be inconsistent with the common belief that ratings can be improved by "dumbing down" course content; however, it should be noted that the instructor's influence on "difficulty" (including behaviors related to "dumbing down") were statistically removed from the measure. An exception was found on progress ratings on "team skills" where the regression weight was negative. It can be inferred that, if the subject matter is inherently very difficult, progress on learning "teamwork" skills may be impeded.

Student academic effort (E_2) made significant contributions to the prediction of 4 of the 15 outcome measures. In three of these instances, its partial regression weight was positive (*developing personal values; acquiring an interest in learning; increasing positive feelings about the field of study*). Classes which enroll a preponderance of students with sound academic work habits and attitudes are likely to be advantaged when students rate their progress on these three criteria. The partial regression weight for this measure was negative on the outcome rating for *gaining a broad, liberal education*. Results for this objective were atypical for both this measure and for that of "Other motivation" (M). Further studies are needed to understand why "frivolous" motivation (M) and academic dedication (E_2) affect progress ratings on this objective in such an unusual manner.

TABLE 24

Statistical Values Derived in Developing "Adjusted" Outcome Measures (Long Form)

Outcome Measure*	Partial Regression Weights					Con-stant	N	R^2
	Enroll	CM	M	D_2	E_2			
21	---	.2737	---	.3102	---	2.5345	27192	.189
22	---	.2426	---	.3204	---	2.5787	27162	.186
23	-.0025	.2953	-.1078	.2445	---	2.5847	23993	.119
24	-.0026	.2994	---	.2408	---	2.6638	19217	.164
25	-.0144	.5544	-.2607	-.1880	---	2.1725	467	.109
26	-.0049	.2117	---	.2569	---	2.8909	9277	.106
27	---	.4644	.1593	.1613	-.1989	1.8951	7282	.224
28	-.0095	.1362	-.1876	---	---	3.5769	14301	.046
29	-.0169	.4453	-.1899	---	---	2.4252	343	.104
30	-.0096	.2493	-.1509	.1949	---	2.7467	1165	.117
31	-.0064	.2764	-.1223	---	.1081	2.7112	780	.077
32	---	.1699	-.2178	.1501	.1669	2.9135	419	.059
40. Pos Att. Twd Field	---	.4699	---	---	.1230	2.3140	35458	.217
41. Tchr Excellent	---	.4573	-.3835	---	---	2.6347	35458	.060
42. Course Excellent	---	.4976	---	---	---	2.2267	35458	.238

*21. Gaining factual knowledge (terminology, classifications, methods, trends)
22. Learning fundamental principles, generalizations, or theories.
23. Learning to *apply* course material (to improve thinking, problem solving, and decisions).
24. Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course.
25. Acquiring skills in working with others as a member of a team.
26. Developing creative capacities (writing, inventing, designing, performing in art, music, drama, etc.).
27. Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature).
28. Developing skill in expressing myself orally or in writing.
29. Learning how to find and use resources for answering questions or solving problems.
30. Developing a clearer understanding of, and commitment to, personal values.
31. Learning to *analyze* and *critically evaluate* ideas, arguments, and points of view.

32. Acquiring an interest in learning more by asking my own questions and seeking answers.

The amount of variance accounted for by these measures of extraneous circumstances, while generally consistent with the literature on the topic, is somewhat lower than expected, given the number of extraneous variables which were considered. Perhaps the major reason is the removal of "teacher effects" from the measures of M, D_2, and E_2. Preliminary studies had shown that, before these effects were removed, measures of enrollment, student motivation, and course difficulty accounted for 30-50 percent of the variance in various outcome measures. But, as has been shown, both student motivation and course difficulty can be powerfully influenced by the instructor. The removal of that influence reduces the relationship of these variables to course outcomes; it also ensures that "adjustments" based on circumstances not under the instructor's control are as accurate and sensitive as possible.

Adjustments to the short form followed the same process as for the long form. However, because the short form contains no items describing teaching techniques or course management decisions, it was possible to employ only three measures of extraneous influence--course motivation (CM), other motivation (M), and Enrollment. The pilot database was used for the new objectives, and the historical database (1994-1995) for those retained from the old form. Results are shown in Table 25.

Findings were consistent with those for the long form. The partial regression weight for class size ("Enrollment") was significant and negative for 9 of the 12 individual progress ratings, but did not make a significant independent contribution to prediction of the three global measures. Course motivation (CM) made a significant contribution to the prediction of all 15 measures of effectiveness. Partial regression coefficients for "Other motivation" (M) were significant, and consistently negative, for 9 of the 12 criterion measures. Therefore, conclusions with respect to the impact of extraneous variables were consistent with those derived from analysis of the long form. However, the measures of course/academic difficulty (D_2) and student academic motivation (E_2), available only on the long form, accounted for a significant amount of criterion variance. Therefore, measures of extraneous variables on the short form were less potent than those employed on the long form. Such measures accounted for an average of 8.8 percent of the variance in outcome measures on the short form, compare to an average of 13.4 percent on the long form.

VI. Reliabilities and Related Statistics

Ratings are never made with perfect reliability. Therefore, when means, standard deviations, or other statistics are reported, they are only estimates. If the same raters completed the IDEA form on another day, somewhat different results would be obtained. This is acknowledged in the IDEA Report to the instructor which displays a range of values**,** rather than a single average, for most measures. This range represents the obtained mean plus and minus one standard error of measurement. Theoretically, if students rated a given teacher an infinite number of times, the odds are two out of three that the "true" average of these ratings would fall in the reported range.

The standard error of measurement reflects two quantities--the degree of variation in the measure from one

class to the next (its standard deviation), and the reliability of the measure.[16] Standard deviations are not affected by the size of the sample, but reliabilities are[17]. The larger the sample, the more reliable the measure.

TABLE 25

Partial Regression Weights and Constants Used in Developing "Adjusted" Outcome Measure (Short Form)

Outcome Measure*	Partial Regression Weights			Constant	N	R^2
	Enroll	CM	M			
1	---	.3209	-.0660	2.8347	27191	.102
2	---	.2369	---	3.0710	27162	.084
3	-.0024	.3139	-.1255	2.8759	23993	.072
4	-.0026	.3016	---	3.0056	19218	.117
5	-.0138	.5377	-.2373	1.9229	467	.100
6	-.0050	.2318	---	3.1836	9277	.070
7	-.0033	.1965	---	3.1233	7282	.034
8	-.0095	.1362	-.1876	3.5768	14301	.046
9	-.0169	.4453	-.1889	2.4252	343	.104
10	-.0092	.2690	-.1406	2.9547	1165	.090
11	-.0064	.2749	-.1003	2.7132	780	.072
12	---	.1613	-.1551	3.1593	919	.011
16. Pos Att. Twd Field	---	.4787	---	2.2851	35458	.212
17. Tchr Excellent	---	.4573	-.3836	2.6347	35458	.060
18. Course Excellent	---	.4976	---	2.2267	35458	.238

* 1. Gaining factual knowledge (terminology, classifications, methods, trends)

[16] The formula is SE_{meas} = (s. d) times.$\sqrt{}(1-r_{11})$.

[17] This is shown by the well known "Spearman-Brown Prophecy Formula", $r^2 = \frac{nr^2}{[1=(n-1)r^2]}$, a specialized case of the "generalizability coefficient" (Crocker & Algina, 1986)

2. Learning fundamental principles, generalizations, or theories.
3. Learning to *apply* course material (to improve thinking, problem solving, and decisions).
4. Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course.
5. Acquiring skills in working with others as a member of a team.
6. Developing creative capacities (writing, inventing, designing, performing in art, music, drama, etc.).
7. Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature).
8. Developing skill in expressing myself orally or in writing.
9. Learning how to find and use resources for answering questions or solving problems.
10. Developing a clearer understanding of, and commitment to, personal values.
11. Learning to *analyze* and *critically evaluate* ideas, arguments, and points of view.
12. Acquiring an interest in learning more by asking my own questions and seeking answers.

Reliabilities of individual items. Sixbury and Cashin (1995) reported reliabilities and standard errors of measurement for all items on the original IDEA form. Separate reports were made for classes with 10, 15, 20, 30, and 40 raters. No re-computations were done for these items. Reliabilities and standard errors of measurement were required, however, for all new items as well as for "scales" based on a combination of items (old and new). All computations were based on classes with 15-34 respondents. The average size for such classes was 22.5, and this figure was used to estimate reliabilities for classes of other sizes (10-14; 15-34; 35-49; 50-99; and 100+).

Reliabilities for individual items were estimated by randomly dividing the students in each class. The means for each half were correlated. This produced an estimate of the reliability for a class of 11 (half the average class size). The Spearman-Brown formula was used to estimate reliabilities for the other class sizes. Results are shown in Table 26.

TABLE 26

Reliabilities and Standard Errors of Measurement for "New" IDEA Items

Item Number		Size of Class									
New	Old	<15		15-34		35-49		50-99		100+	
		r_{11}	SE_M	r_{11}	SE_M	r_{11}	SE_M	r_{11}	SE_M	r_{11}	SE_M
1	60	.58	.39	.71	.32	.81	.26	.88	.21	.92	.17
3	64	.49	.39	.62	.34	.73	.29	.82	.23	.88	.19
5	61	.85	.42	.92	.31	.95	.24	.97	.18	.98	.14
9	67	.81	.30	.89	.23	.93	.17	.96	.13	.98	.11
14	62	.75	.42	.85	.33	.91	.26	.95	.20	.97	.16
15	68	.74	.33	.84	.25	.90	.20	.94	.15	.96	.12
16	72	.79	.33	.87	.25	.92	.19	.96	.15	.97	.12
17	69	.78	.30	.87	.23	.92	.18	.95	.14	.97	.11
18	71	.78	.34	.87	.26	.92	.20	.95	.15	.97	.12
20	70	.79	.39	.88	.30	.93	.23	.96	.17	.97	.14
25	50	.84	.36	.91	.27	,95	.20	.97	.16	.98	.12
29	51	.82	.36	.90	.27	.94	.21	.96	.16	.98	.12
30	53	.62	.35	.75	.29	.83	.23	.90	.18	.93	.15

31	55	.62	.36	.74	.30	.83	.24	.90	.19	.93	.15
32	54	.65	.37	.77	.30	.85	.24	.91	.19	.94	.15
40	38	.71	.35	.82	.28	.88	.22	.93	.17	.96	.13
41	44	.79	.31	.88	.23	.93	.18	.96	.14	.97	.11
42	45	.73	.32	.84	.25	.90	.20	.94	.15	.96	.12

Reliability is adversely affected if the range of ratings is constricted; such ranges also tend to reduce the standard error of measurement. Therefore, an item with a relatively high reliability will not necessary have a low standard error of measurement, as seen in Table 26.

In general, the standard error of measurement tended to the lowest for the "Global outcomes" measures (new items 40-42), averaging .25 for classes in the 15-34 enrollment range. They were more variable for the methods items (items 1-20) than for the items assessing progress on objectives (items 25-32), but the averages for both sets were about the same (.28 and .29, respectively, for class size 15-34).

These estimates are a little larger than those reported by Sixbury and Cashin for the items on the original IDEA form. For a sample of 500 classes enrolling exactly 20 students, the standard errors of measurement for the 20 methods items averaged .247; for the 10 ratings of progress on specific objectives, they averaged .249. From these results, we anticipate that the standard errors for the new items may be somewhat over-estimated from the pilot results. Of course, as a national database is developed, new and more stable estimates will be made.

Reliabilities of More Complex Measures. As noted earlier, the 20 methods items included on the new form (10 from the original form; 10 from the pilot items) were grouped into five scales representing dimensions of teaching styles: *Student-Faculty Contact* (new form Items 1, 2, and 20); *Involving Students* (new form Items 5, 9, 14, 16, and 18); *Establishing Expectations* (new form Items 3, 4, 8, 13, and 15); *Clarity of Communication* (new form Items 6, 10, and 11); and *Assessment-Feedback* (new form Items 7, 12, 17, and 19). The reliability of scores derived by averaging ratings on these items were computed using data from classes enrolling 15-34 students in the pilot sample. Cronbach's Coefficient Alpha was computed for each scale; results were used as to estimate scale reliability for classes averaging 22.5. The Spearman-Brown formula was then employed to estimate reliabilities for classes enrolling 10-15, 35-49, 50-99, and 100+. These figures, together with the standard deviations calculated from classes enrolling 15-34, were used to estimate standard errors of estimate for each size range. Results are shown in Table 27.

TABLE 27

Reliability and Standard Error of Measurement for Methods Scales

Scale	Size of Class									
	10-15		15-34		35-49		50-99		100+	
	r_{11}	SE_M	r_{11}	SE_M	r_{11}	SE_M	r_{11}	SE_M	r_{11}	SE_M
S-Fac. Contact	.75	.21	.85	.16	.90	.13	.94	.10	.96	.08
Involving Students	.78	.21	.88	.16	.92	.13	.95	.10	.97	.08
Estab. Expectations	.83	.20	.91	.15	.94	.12	.97	.09	.98	.07

Clarity Commun.	.78	.21	.88	.16	.92	.13	.95	.10	.97	.08
Assessment/Feedback	.63	.23	.76	.19	.84	.16	.90	.12	.93	.10

As expected, reliabilities were higher, and standard errors lower, for these combinations of items than for the individual items reported in Table 26. These results suggest that the items within the scales were internally consistent, offering reason to conduct further research to establish the meaning and usefulness of the scales. One other important outcome included on the IDEA Report to the Instructor is that for *Progress on Relevant Objectives*. This measure considers the student progress ratings on objectives chosen by the instructor as "Important" or "Essential." To construct it, the average progress rating on each selected objective was first converted to a "T Score," a standardized score which not only makes it easy to compare a given rating with those for others choosing the objective as "Important" or "Essential" but also puts all progress ratings on a scale with the same mean (50) and standard deviation (10). In computing an average rating for *Progress on Relevant Objectives*, standard scores for "Essential" ratings were given a double weight while those for "Important" objectives received a single weight (and those for objectives not chosen were omitted; i. e., given no weight).

Using data from the pilot sample, reliabilities were computed for both raw and "adjusted" scores for classes enrolling 13-17, 22-29, and 35-49. To do this, students in each class were randomly divided, and the average rating for each of these two halves were correlated. The Spearman-Brown formula was applied to estimate the reliability for classes enrolling an average of 15, 25, and 42 students. It was also used to project reliabilities for classes of 10-14, 15-34, 35-49, 50-99 and 100 or over. Table 28 summarizes these reliabilities and the associated standard errors of measurement.

As expected, the reliability of this summary measure was generally higher than those for individual progress ratings (Table 26). Standard errors are not comparable, since those in Table 28 are presented in T score, rather than raw score, units. Estimates for adjusted ratings were somewhat less accurate than for raw ratings, reflecting the lack of perfect reliability of measures used to make adjustments. The largest proportion of classes participating in the IDEA program is in the 15-34 range; for such classes, standard errors on this measure were 3.2 and 3.6 T score points for raw and adjusted ratings, respectively.

Table 28

Reliability and Standard Error of Measurement for "Progress on Relevant Objectives" (Raw and Adjusted)

Size of Class	Raw Progress Rating		Adjusted Progress Rating	
	r_{11}	SE_M	r_{11}	SE_M
10-14	.78	4.5	.74	5.1
15-34	.87	3.2	.85	3.6
35-49	.92	2.5	.90	2.7
50-99	.95	1.9	.94	2.2
100+	.97	1.6	.96	1.8

Inferring "Strengths" and "Weaknesses". The IDEA system attempts to provide diagnostic assistance to its

users by identifying relevant "strengths" and "weaknesses" in instructional approaches. The first step in this process is to identify those methods items that make an independent contribution to the prediction of the progress rating on a given objective. These are called "relevant methods." If progress ratings on an objective chosen as "important" or "essential" by the instructor were in the "average or below" range, then his/her average rating on relevant methods is compared with that for "similar" courses. A "weakness" is defined as a rating on a "relevant" item that is at least 0.3 (approximately one standard error) below that for similar courses. Similarly, a "strength" is a relevant item whose average rating exceeds that of similar courses by 0.3 or more.

"Similar" courses have been defined as those in the same size range (Small=less than 15; Medium=15-34; and Large=35 or more[18]) and in the same level of motivation, defined as the average rating of "I had a strong desire to take this course" (I=below 3.0; II=3.0-3.4; III=3.5-3.9; IV=4.0-4.4; V=4.5+). Although this measure of motivation may be flawed, as noted earlier, it was used in the analyses reported below because there is not yet an adequate database for more sophisticated measures.

To identify "strengths" and "weaknesses," it is necessary to determine the mean rating of each of the 20 methods items for each of 15 groups (3 sizes multiplied by 5 motivation levels). For the ten items which were retained from the old form, the 1994-1995 database was employed. Results are reported in Table 29.

TABLE 29

Means for Methods Items Retained from Original IDEA Form
For Groups Defined by Size and Motivation Level

Item	Motivation Level	Size		
		Small	Medium	Large
2	I	3.74	3.72	3.46
	II	3.90	3.87	3.72
	III	4.03	4.00	3.86
	IV	4.18	4.13	4.06
	V	4.34	4.28	4.17
4	I	4.03	4.04	3.95
	II	4.20	4.21	4.18
	III	4.33	4.35	4.33
	IV	4.47	4.49	4.51
	V	4.62	4.61	4.55
6	I	3.84	3.85	3.78
	II	4.02	4.04	4.02
	III	4.16	4.18	4.16
	IV	4.30	4.32	4.33
	V	4.47	4.45	4.39
7	I	3.33	3.31	3.09

[18]For some purposes, the IDEA system has divided classes enrolling more than 34 students into "Large" (35-99) and "Very large" (100+). However, even with a database of over 35,000 classes, there were two few "Very large" classes to justify the use of all four size categories in this analysis.

	II	3.43	3.41	3.24
	III	3.53	3.50	3.32
	IV	3.68	3.62	3.49
	V	3.90	3.80	3.71

TABLE 29, Concluded

Item	Motivation Level	Size		
		Small	Medium	Large
8	I	3.42	3.38	3.23
	II	3.54	3.54	3.45
	III	3.71	3.68	3.57
	IV	3.86	3.82	3.79
	V	4.05	4.01	3.96
10	I	3.73	3.82	3.65
	II	3.94	4.00	3.96
	III	4.10	4.14	4.12
	IV	4.25	4.27	4.32
	V	4.45	4.43	4.33
11	I	3.84	3.83	3.87
	II	4.08	4.08	4.17
	III	4.24	4.27	4.34
	IV	4.39	4.41	4.52
	V	4.52	4.51	4.50
12	I	4.01	4.04	3.97
	II	4.14	4.18	4.16
	III	4.29	4.30	4.29
	IV	4.38	4.40	4.38
	V	4.48	4.47	4.38
13	I	3.51	3.50	3.43
	II	3.74	3.77	3.78
	III	3.96	3.97	3.99
	IV	4.17	4.17	4.24
	V	4.38	4.34	4.25
19	I	3.61	3.61	3.26
	II	3.77	3.74	3.49
	III	3.91	3.85	3.61
	IV	4.04	3.96	3.76
	V	4.18	4.06	3.87

As was noted earlier, the means for most items decrease as class size increases. This trend is less sharp than that for motivation level; the means for Motivation Level V are typically about 0.6 above those for Level I, equivalent to about two standard errors of measurement.

For the ten new items, the pilot database was far too small to provide the required empirical data. Still, it was important to classify relevant items as "strengths" or "weaknesses." A two-part strategy was employed to

arrive at estimates which could be used for that purpose.

First, classes were sorted by size (10-14+Small; 15-34=Medium; and 35-99=Large) and means were computed for each item. Second, medium-sized classes were sorted by motivation level, using the same definition as that employed for items retained from the original IDEA form. The differences between levels found for these medium sized classes were assumed to be representative of differences between motivation levels for classes of the other two sizes. Using this assumption, it was possible to use data regarding differences among sizes to estimate means for each of the 15 size x level of motivation categories. Results are shown in Table 30.

TABLE 30

Statistics Used in Estimating Means for New Methods Items For Groups Defined by Size and Motivation Level

Item N	Size			Motivation Level, Medium Sized Classes				
	Small	Medium	Large	I=Low	II	III	IV	V=High
1	4.03	3.87	3.55	3.66	3.81	3.94	4.06	4.22
N	181	801	130	179	223	176	138	85
3	4.06	3.96	3.78	3.76	3.86	3.99	4.12	4.27
N	181	800	128	179	223	175	137	86
5	3.35	3.25	2.82	3.40	3.30	3.18	2.99	3.36
N	181	804	129	180	223	176	139	86
9	3.80	3.76	3.50	3.84	3.76	3.60	3.55	3.50
N	47	255	59	121	59	41	28	6
14	3.81	3.55	3.39	3.30	3.39	3.46	3.66	4.25
N	181	803	129	180	223	176	138	86
15	3.95	3.71	3.46	3.64	3.72	3.80	3.91	4.00
N	47	255	59	121	59	41	28	6
16	4.07	3.85	3.62	3.51	3.68	3.85	4.04	4.14
N	113	481	184	91	118	136	104	32
17	4.09	4.00	3.82	4.00	4.00	4.00	4.15	4.20
N	47	255	59	121	59	41	38	6
18	3.89	3.79	3.49	3.75	3.78	3.85	3.96	4.05
N	47	255	59	121	59	41	28	6
20	3.84	3.71	3.50	3.57	3.67	3.78	3.85	4.13
N	140	528	201	99	132	143	122	32

To illustrate the process, consider the results for new Item 3. The first part of Table 30 shows that the average for small classes was .10 above that for medium-sized classes (4.06-3.96), and the average for the latter was .18 above that for large classes (3.96-3.78). In the right hand section of Table 30, it can be seen that the mean for Item 3 steadily increased from motivation Level I (3.76) through Level V (4.27). Difference between Level I and II (3.86-3.76=.10), Level II and III (3.99-3.86=.13), III and IV (4.12-3.99=.13), and IV and V (4.27-4.12=.15) were assumed to hold for the other two size groups. Therefore, for

small classes, the estimated means for Levels I through V were 3.86, 3.96, 4.09, 4.22, and 4.37.[19]

VII. Summary

Because the IDEA system has been unrevised for the past 23 years, there is reason to believe that the content of the form may be in need of serious revision. Following evolutionary (and revolutionary) changes in society, higher education has responded with new ideas about purposes and new approaches for addressing them. IDEA Center staff, as well as many of its advisors, believed that these changes were sufficient to justify a thorough revision of the IDEA instrument.

Through extensive consultation with experienced users of the IDEA system and with authorities in the evaluation of instruction and faculty development, a number of instructional objectives not included on the current IDEA form were identified. Instructional techniques recommended by contemporary leaders in instructional methodology were also identified.

Items were written to describe a total of 6 "new" objectives and 15 "new" methods. These were pilot tested as "extra questions" by 10 institutions who voluntarily participated in a tryout experiment. Although not randomly selected, basic IDEA results for these institutions were very similar to those for the entire IDEA database, suggesting that their results on the 21 proposed new items would be representative of those for the user population.

Results of this "tryout" experiment showed that five of the six new objectives were highly pertinent to the description of instructional purposes. Items describing these five objectives were selected for inclusion on the revised form. Three items describing objectives which were on the original IDEA form were excluded from the revised form, either because they were infrequently used or because their content overlapped that of other objectives chosen for the revision. As a result, the new form contains 12, rather than 10, statements of objectives.

To select "methods" items which would have the most diagnostic utility, multiple regression was employed. Two data sets were used. Approximately 35,000 classes from the 1994 and 1995 years were used to relate the existing 20 methods items to progress ratings on each of the seven objectives which were retained from the original IDEA form. Approximately 3500 classes from the pilot institutions were used to related these 20 items to progress ratings on the five "new" objectives. These classes were also used to relate the 15 proposed methods items to (a) the seven existing objectives and (b) the five new objectives.

All 35 items had some diagnostic utility. But since it had been decided that the revised IDEA form should contain no more than 20 methods items, it was necessary to exclude 15 of them. Eight exclusions were made purely on the basis of utility in predicting progress ratings on the 12 objectives. The other seven considered (a) the adequacy with which chosen items reflected important dimensions of instruction and (b) the number of items which had diagnostic value for improving effectiveness in achieving a given instructional objective.

In addition to changes in content, it was intended that the revised IDEA system should improve the way in which "extraneous circumstances" (conditions beyond the instructor's control) could be taken into account. Five such measures were employed: (1) size of enrollment; (2) course motivation (average response to "I really wanted to take the course regardless of who taught it"); (3) other motivation [a residual which adjusts the average response to one item ("I had a strong desire to take this course") in terms of the average response to another item ("I really wanted to take a course from this instructor")]; (4) discipline-related difficulty, the portion of the rating of "difficulty" which could not be traced to teacher behavior; and (5) student academic

[19] These estimates are highly tentative and will be replaced by empirical results as they become available.

effort, the portion of the rating of student effort which could not be accounted for by teacher methods. Each of these five measures accounted for some of the variation in at least one of the outcome measures (either global measures or ratings of progress on specific objectives). Appropriate "adjustments" were made in these measures of outcomes (effectiveness) so that instructors would be neither advantaged nor disadvantaged by the influence of factors beyond their control.

To prepare a useful report to the faculty member, basic statistics were computed related to reliability and to the effect of class size and motivation level on specific methods. Some of this information was available from IDEA's historic database of over 100,000 classes. But data regarding new items was inadequate for making stable estimates of these statistical values. As the database for the revised IDEA form expands, these estimates will be replaced by more stable empirical findings.

References

Braskamp, L. A. and Ory, J. C. (1994) *Assessing Faculty Work*. San Francisco: Jossey-Bass, 353 pp.

Centra, J. A. (1993). *Reflective Faculty Evaluation*. San Francisco: Jossey-Bass, 244 pp.

Chickering, A. W. and Gamson, Z, 1987. Seven principles of good practice in undergraduate education. *American Association for Higher Education Bulletin, 39,* 3-7.

Feldman, K. A. (1976) Grades and college students' evaluations of their courses and teachers. *Research in Higher Education, 4,* 69-111.

Feldman, K. A. (1978) Course characteristics and college students' ratings of their teachers: what we know and what we don't know. *Research in Higher Education, 9,* 199-242.

Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education, 10,* 149-172.

Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education, 18,* 3-124.

Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: a closer look. *Research in Higher Education, 21,* 45-116.

Horst, P. (1941). *Prediction of Personal Adjustment*. Bulletin #48. New York: Social Science Research Council.

Hoyt, D. P. (1973). Measurement of instructional effectiveness. *Research in Higher Education, 1,* 367-378.

Sixbury, G. R. and Cashin, W E. (1995). Description of database for the IDEA diagnostic form. *Technical Report No. 9*, Center for Faculty Evaluation and Development. Manhattan, Kansas: Kansas State University.

APPENDIX

RELATIONSHIP BETWEEN "OLD" AND "NEW" IDEA ITEM NUMBERS

Item Number		
Old	New	Item
1	-	
2	2	Found ways to help students answer their own questions.
3	-	
4	-	
5	-	
6	-	
7	-	
8	4	Demonstrated the importance and significance of the subject matter.
9	-	
10	6	Made it clear how each topic fit into the course.
11	7	Explained the reasons for criticisms of students' academic performance.
12	-	
13	-	
14	-	
15	8	Stimulated students to intellectual effort beyond that required by most courses.
16	-	
17	10	Explained course material clearly, and explanations were to the point.
18	11	Related course material to real life situations.
19	-	
20	13	Introduced stimulating ideas about the subject.
A(40)	12	Gave tests, projects, etc. that covered the most important points of the course.
B(41)	19	Gave projects, tests, or assignments that required original or creative thinking.
(60)	1	Displayed a personal interest in me and my learning.
(64)	3	Scheduled course work (class activities, tests, projects) in ways which encouraged students to stay up to date in their work.
(61)	5	Formed "teams" or "discussion groups" to facilitate learning.
(65)	9	Encouraged students to use multiple resources (e. g., data banks, library holdings, outside experts) to improve understanding.
(62)	14	Involved students in "hands on" projects such as research, case studies, or "real life" activities.
(68)	15	Inspired students to set and achieve goals which really challenged them.
(72)	16	Asked students to share ideas and experiences with others whose backgrounds and viewpoints differ from their own.
(69)	17	Provided timely and frequent feedback on tests, reports, projects, etc. to help students improve.
(66)	18	Asked students to help each other understand ideas or concepts.
(70)	20	Encouraged student-faculty interaction outside of class (office visits, phone calls, e-mail, etc.)

(Continued, next page)

Item Number		
Old	New	Item
21	21	Gaining factual knowledge (terminology, classifications, methods, trends).
22	22	Learning fundamental principles, generalizations, or theories.
23	23	Learning to *apply* course material to improve rational thinking, problem-solving, and decision making.
24	24	Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course.
25	-	
26	26	Developing creative capacities (writing, inventing, designing, performing in art, music, drama, etc.)
27	-	
28	27	Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature, etc.)
29	28	Developing skill in expressing myself orally or in writing.
30	-	
(50)	25	Acquiring skills in working with others as a member of a team.
(51)	29	Learning how to find and use resources for answering questions or solving problems.
(53)	30	Developing a clearer understanding of, and commitment to, personal values.
(55)	31	Learning to *analyze* and *critically evaluate* ideas, arguments, and points of view.
(54)	32	Acquiring interest in learning more by asking my own questions and seeking answers.
31	33	Amount of reading.
32	34	Amount of work in other (non-reading) assignments.
33	35	Difficulty of subject matter.
34	-	
35	37	I worked harder on this course than on most courses I have taken.
36	36	I had a strong desire to take this course.
38	40	As a result of taking this course, I have more positive feelings toward this field of study.
39	-	
C(42)	38	I really wanted to take a course from this instructor.
D(43)	39	I really wanted to take this course regardless of who taught it.
E(44)	41	Overall, I rate this instructor an excellent teacher.
F(45)	42	Overall, I rate this course as excellent.
G(46)	-	
	A	As a rule, I put forth more effort than other students on academic work.
	B.	The instructor used a variety of methods--not only tests--to evaluate student progress on course objectives.
	C	The instructor expected students to take their share of responsibility for learning.
	D	The instructor had high achievement standards in this class.
	E	The instructor used educational technology (e. g., Internet, E-mail, computer exercises, multi-media presentations, etc.) To promote learning.