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.