The Relationship of Test Performance to ISDP Rating in Organic Chemistry Texts

Soo-Young Choi
Instructional Science Department
Brigham Young University
Provo, UT 84602

M. David Merrill
University of Southern California
Los Angeles, CA 90007

Ed Callahan
Courseware, Inc.
10075 Carroll Canyon Road
San Diego, CA 93131

Richard T. Hawkins
Chemistry Department
Brigham Young University
Provo, UT 84602

Robert F. Norton
Center of Evaluation and Management
Kuwait University
Box 5969 Kuwait

Introduction

The Chemistry Department at Brigham Young University (BYU) has used two different textbooks for the beginning organic chemistry course. The two textbooks are *Introduction to Organic Chemistry* by Streitwieser and Heathcock (1976) and *Organic Chemistry* by Morrison and Boyd (1973). The two different textbooks were used to determine the most appropriate textbook for the organic chemistry course.

For the evaluation of students' performance, a national standardized organic chemistry test was used. Students' scores on the test represented student competency in this area. Therefore, this study used Form 1974 American Chemical Security (ACS) Cooperative Exam in organic chemistry, prepared by the Examination Committee of the Division of Chemical Education of the American Chemical Security through its Organic Chemistry Subcommittee. The test has 70 multiple-choice items.

Prior to this study the Chemistry Department at BYU had selected a textbook based upon student performance (ACS organic chemistry test) and intuition; however, no empirical study such as the one described in this article had been conducted to assist with the selection process. Therefore, this study was designed to find out which textbook is "better" from an instructional science perspective and to provide data bearing on the Chemistry Department's textbook selection decision.

Merrill, Olsen, and Coldeway (1976) have developed an analytic tool to evaluate instructional materials. This tool is called the Instructional Strategy Diagnostic Profile (ISDP) (Note 1). The ISDP is an instrument to judge the consistency and adequacy of a content/task objective, presentation, and evaluation.

> ". . . this study was designed to find out which textbook is 'better' from an instructional science perspective. . . ."

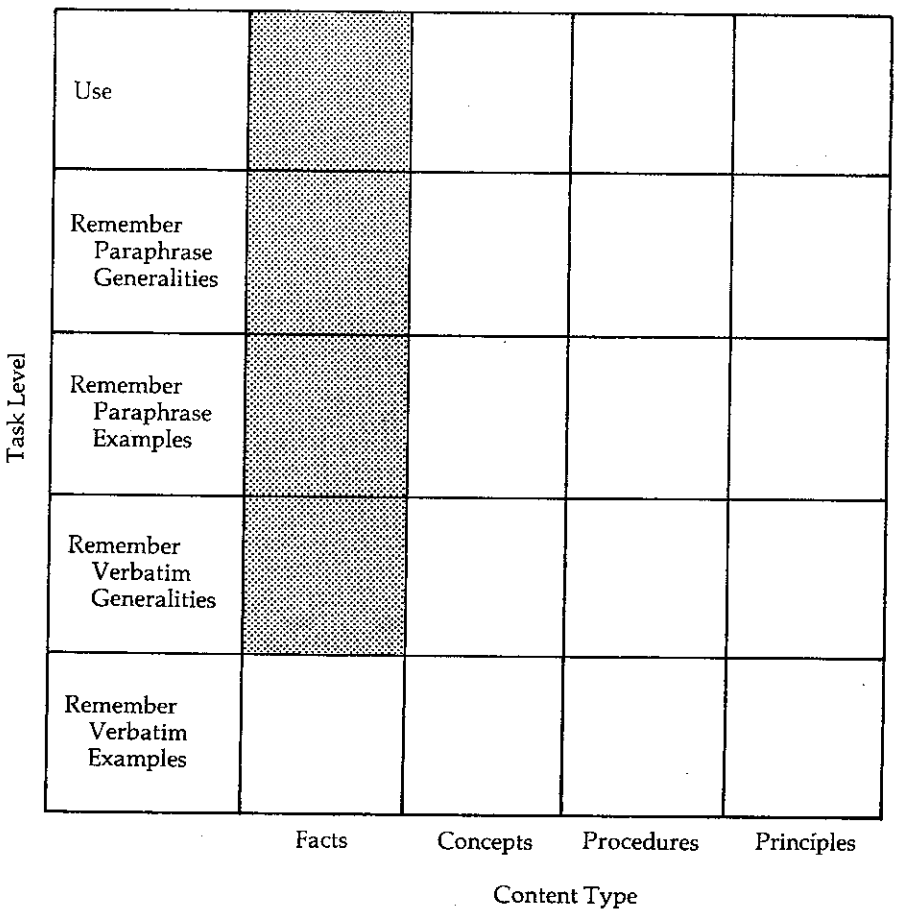
Overview of the Instructional Strategy Diagnostic Profile[1]

The ISDP is an analytic tool, in the form of a set of prescriptions, that facilitates the evaluation and revision of existing instruction and the design of new instruction. The ISDP is designed to evaluate instructional materials on two main criteria: consistency and adequacy. Consistency must be determined before adequacy can be assessed.

The consistency criterion is met if the instructional objectives, test items, and the instructional presentation are consistent. Determining consistency is accomplished in two steps. First, the instructional objectives and test items are classified on two dimensions: (a) the performance, or task level, required of the student and (b) the type of instructional content. These two dimensions then are combined to form a task/content classification matrix, which is used to classify objectives, test items, and instructional presentation. This matrix is illustrated in Figure 1. If an objective and its corresponding test item can be classified in the same cell of the matrix, they are considered to be consistent. The second step involved rating the consistency between instruction and objective/test items. The ISDP requires that different components of instructional presentation, called primary presentation forms, be present for different combinations of task level and content type. If the combination of primary presentation forms required for the task level and content type of each objective/test item is present, then the instruction is consistent with the objective/test items.

Once it has been determined that instructional materials are consistent, the adequacy criterion is assessed. This is done by determining whether or not the instructional presentation adequately communicates the "to-be-learned" information. Variables that are hypothesized as affecting instructional ade-

[1]This information is taken verbatim from a report by M. D. Merrill, R. E. Richards, R. V. Schmidt, and N. D. Wood entitled *Empirical Validation of Selected Instructional Strategy Diagnostic Profile Prescriptions* (NPRDC Tech. Rep. 77-43, September 1977) pages 1 through 5.



Note: Facts can be tested only at the *Remember Verbatim* level.

FIGURE 1. ISDP task/content classification matrix.

quacy during ISDP development include the following: (a) isolation (i.e., is the relevant information separated and clearly identified?), (b) help (i.e., is explanatory or nmemonic information provided?), and (c) matching (i.e., are the examples and practice items matched?). Thus, instruction is based on these variables to obtain an adequacy index. Each primary presentation form within the instruction may be rated as more or less adequate.

In the following sections, the task/content matrix will be described in greater detail, the primary presentation forms will be discussed, and validation experiments testing the consistency and adequacy assumptions will be presented.

Task/Content Classification Matrix

As shown in Figure 1, the task dimension of the task/content classification matrix is comprised of several levels, the broadest of which consists of the strategies *Use* and *Remember. Use* is defined as the act of applying a general relationship to a specific situation where it has not been previously applied, and *remember,* as the act of bringing to mind something that has been previously encountered. Thus, a use item (or objective) would require the student to respond by applying a generality to a newly encountered example; that is, one that has not been previously displayed to the student as part of the instructional presentation. A remember test item (or objective) would require the student to respond by recognizing or recalling a generality or example that has been previously encountered. *Generality* is defined as a statement of definition or relationship that can be applied to more than one specific object or event; and *example,* as a specific object or event or its representation that does or could exist in the real world.

The use level cannot be divided into sublevels—it always requires newly encountered examples. The reason for this is obvious—if an example had been previously encountered, the test item or objective would be classified at the remember level. The remember level, however, can be divided into sublevels, either paraphrase or verbatim. *Paraphrase* means equivalent in meaning but expressed in other words; *verbatim,* word-for-word or exactly the same. Thus, a *paraphrase generality* means that synonyms have been substituted for the substantive words (nouns, verbs, and modifiers) of the original statements; and *paraphrase example* means that the same object or event is presented but that the form or representation used to exhibit this object or event has been modified. A *verbatim generality/example* requires the student to recognize or restate the same words that were used previously to present the generality/example. All paraphrase and verbatim generalities and examples have been previously encountered by the student.

As shown in Figure 1, the content dimension of the matrix involves four mutually exclusive content categories: facts, concepts, procedures, and principles. Except for facts, for which there can be no generalities, all types can be tested at any of the task levels. These categories are defined as follows:

A *fact* is a one-to-one association of a symbol and a specific object or event.

A *concept* is a class of objects, events, or symbols that (a) share critical attributes, (b) can be referenced by a name or symbol, and (c) have discriminately different individual members.

A *procedure* is a series of steps required to produce an example of an outcome class. Each step may involve the temporal or spatial ordering of specific objects, events, or symbols or a branching decision, based either on a fact or the classification of an example of a concept. A procedure is often characterized as "how to do something."

A *principle* is a predictive relationship between specific examples of a concept, or among a set of related concepts, which explains why an example of a particular class is produced as a result of a particular manipulation.

Further information concerning these content categories is provided in Merrill (Note 1), which was designed to teach users how to classify test items and objectives according to the ISDP task/content classification matrix. (It should be noted that some of the variables

identified as affecting instructional adequacy are not applicable to all content types and task levels.)

Primary Presentation Forms

The ISDP defines the instructional presentation form or display as the fundamental unit of instructional strategy. As indicated previously, the instructional presentation forms must meet consistency and adequacy requirements.

Four primary presentation forms or displays, which represent the various ways that information can be presented, have been defined (Merrill, Note 1):

1. Tell via generality (TG)—A display that presents a definition of a concept, an algorithm that describes a procedure, or a proposition that expresses a principle.
2. Tel via example (Teg)—A display that illustrates how a generality applies to a specific example.
3. Question via generality (QG)—Not used in this research study.
4. Question via example (Qeg)—A display that presents an example and requires the student to respond to the example *or* it presents a name or generality and requires the student to respond by providing an example.

Hypotheses

The hypotheses of this study are:

1. There is a significant difference between ISDP ratings for the two texts.
2. There is a significant difference on test performance between students using the two different texts.
3. There is a significant linear relationship between ISDP ratings and test performance of the two texts.
4. There is a significant difference between the text selected by the Chemistry Department and the text with the lower ISDP rating and lower performance.

Methods

Subjects

Four undergraduate organic chemistry classes at BYU were selected for this study. Two of the classes used *Introduction to Organic Chemistry* (by Streitwieser and Heathcock) and two of the classes used the *Organic Chemistry* text (by Morrison and Boyd).

Materials

The materials of this study consist of two organic chemistry texts used at BYU: *Introduction to Organic Chemistry* and *Organic Chemistry* mentioned in the preceding paragraph.

Design

Student posttest scores were obtained on the National Standard Organic Chemistry test, the 1974 ACS Cooperative Exam in organic chemistry. A discrimination analysis was conducted on each item of the test comparing test item scores for those students who used the *Introduction to Organic Chemistry* text with test item scores for those students who used *Organic Chemistry*.[2] A discrimination index was arrived at using the following formula: D = (number of incorrect responses for each textbook)/(total number of respondents). Specific test items were selected for analysis if they had a percentage difference greater than 11 percent between the discrimination indices. A list of 18 items of the total possible 70 items of the posttest were selected for analysis. (The test items were numbers 2, 13, 17, 18, 27, 29, 33, 40, 43, 48, 54, 56, 57, 62, 66, 68, and 69 of the Form 1974 ACS Cooperative Exam in organic chemistry.)

An ISDP was determined for each posttest item. A systematic procedure was followed for the determination of each item's profile. The following is the step-by-step procedure used:

Step 1. Task level of item was determined.

Step 2. Content type of item was determined.

Step 3. The correct ISDP form was selected for the item dependent upon the task/content of the item. (See Figures 2 to 5.)

Step 4. Both texts were surveyed for location of presentation of information relevant to test item.

Step 5. Presentation-test consistency was determined (i.e., presence or absence and number of prescribed components for each test item).

Step 6. Presentation adequacy was determined (i.e., degree to which presentation components meet ISDP criteria).

When all of the selection posttest items were profiled according to the ISDP, a cumulative rating of overall text consistency and overall text adequacy was computed for each item.

Statistical analyses were conducted on the data to address the stated hypotheses of the study. The following are the specific analyses conducted:

1. A *t* test to compare ISDP consistency and adequacy cumulative ratings of the two texts.
2. A *t* test to compare posttest scores of the students using the two texts.
3. A linear regression of ISDP cumulative ratings on posttest scores.
4. Another comparison was made between the text selection of the Chemistry Department with the findings from the ISDP.

Results

Hypothesis 1 stated: "There is a significant difference between ISDP ratings for the two texts." The results of this study support the above hypothesis. The Morrison and Boyd text received higher ISDP ratings than the Streitwieser and Heathcock text (t = 3.104,

> ". . . the text with the higher ISDP ratings . . . also produced higher student performance."

> ". . . we suggest that the present materials be reassessed using the complexity of performance task measure. . . ."

[2]This discrimination analysis was originally conducted using data from all available sources using these texts.

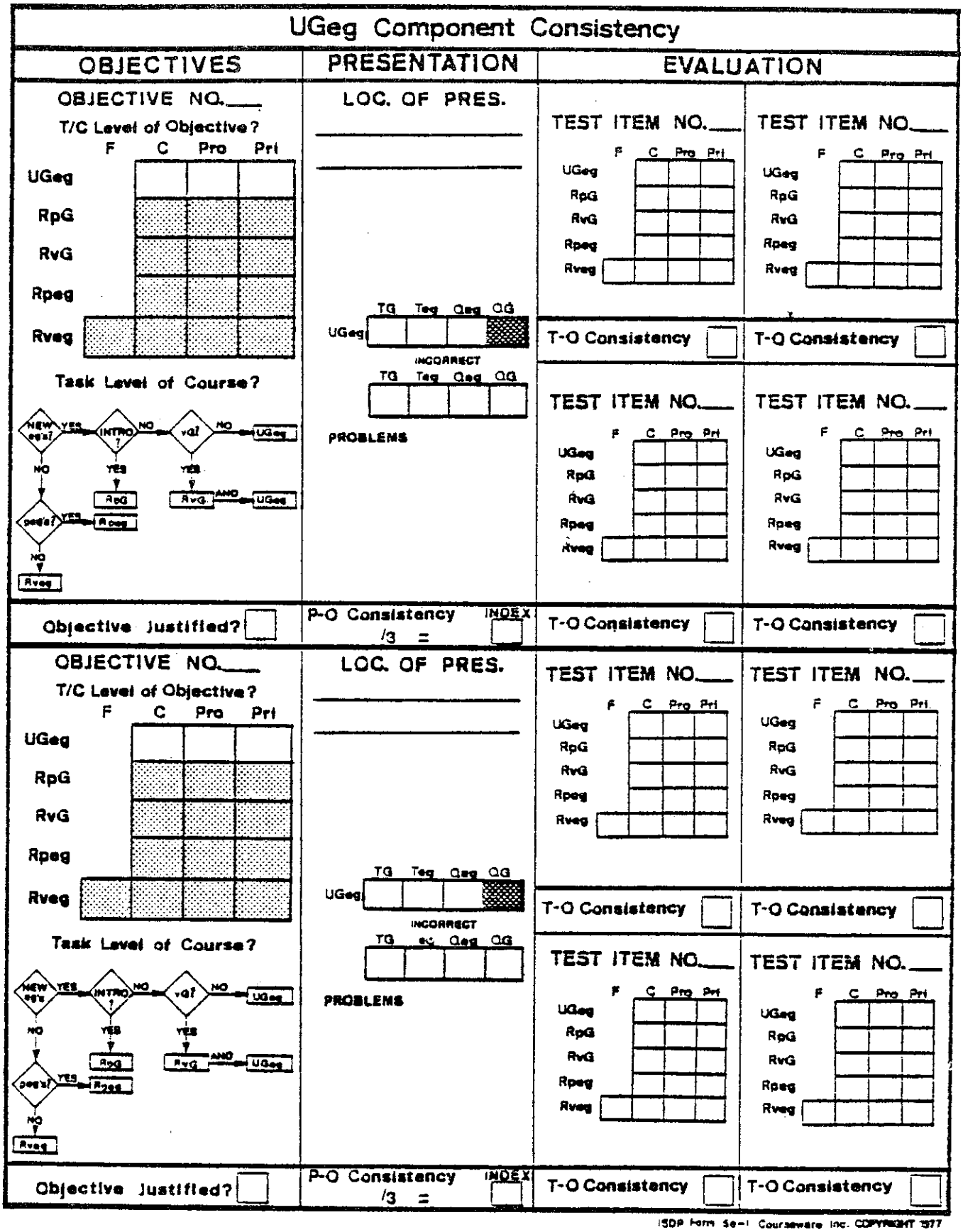
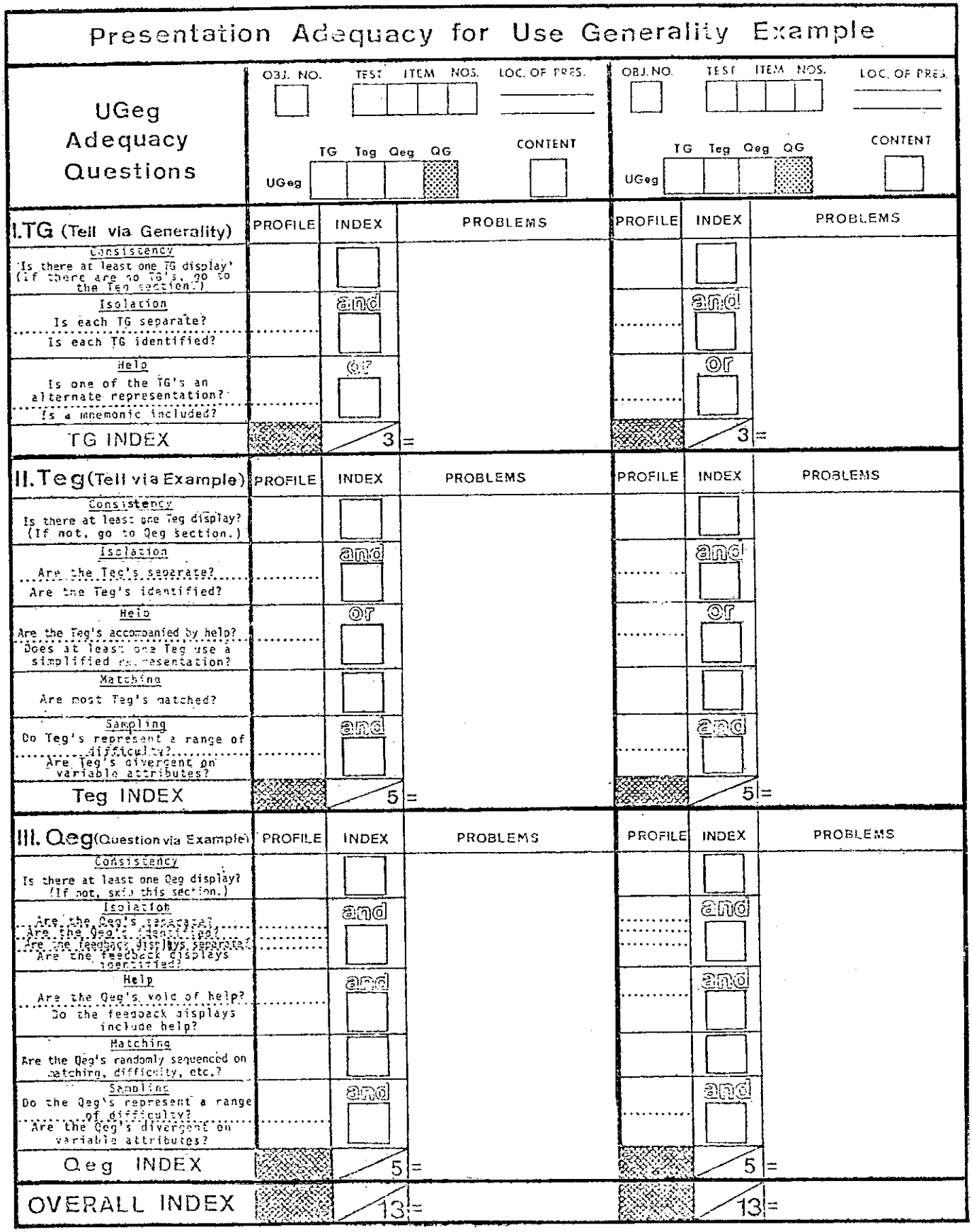


FIGURE 2. Sample ISDP form: UGeg component consistency.



Presentation Adequacy for Use Generality Example

UGeg Adequacy Questions

OBJ. NO. | TEST ITEM NOS. | LOC. OF PRES. | UGeg TG Teg Qeg QG | CONTENT

OBJ. NO. | TEST ITEM NOS. | LOC. OF PRES. | UGeg TG Teg Qeg QG | CONTENT

I. TG (Tell via Generality)

PROFILE | INDEX | PROBLEMS

Consistency
Is there at least one TG display? (If there are no TG's, go to the Teg section.)

Isolation
Is each TG separate?
Is each TG identified?

Help
Is one of the TG's an alternate representation?
Is a mnemonic included?

TG INDEX ___/3 =

II. Teg (Tell via Example)

PROFILE | INDEX | PROBLEMS

Consistency
Is there at least one Teg display? (If not, go to Qeg section.)

Isolation
Are the Teg's separate?
Are the Teg's identified?

Help
Are the Teg's accompanied by help?
Does at least one Teg use a simplified representation?

Matching
Are most Teg's matched?

Sampling
Do Teg's represent a range of difficulty?
Are Teg's divergent on variable attributes?

Teg INDEX ___/5 =

III. Qeg (Question via Example)

PROFILE | INDEX | PROBLEMS

Consistency
Is there at least one Qeg display? (If not, skip this section.)

Isolation
Are the Qeg's separate?
Are the Qeg's identified?
Are the feedback displays separate?
Are the feedback displays identified?

Help
Are the Qeg's void of help?
Do the feedback displays include help?

Matching
Are the Qeg's randomly sequenced on matching, difficulty, etc.?

Sampling
Do the Qeg's represent a range of difficulty?
Are the Qeg's divergent on variable attributes?

Qeg INDEX ___/5 =

OVERALL INDEX ___/13 =

FIGURE 3. Sample ISDP form: Presentation adequacy for *use generality* example.

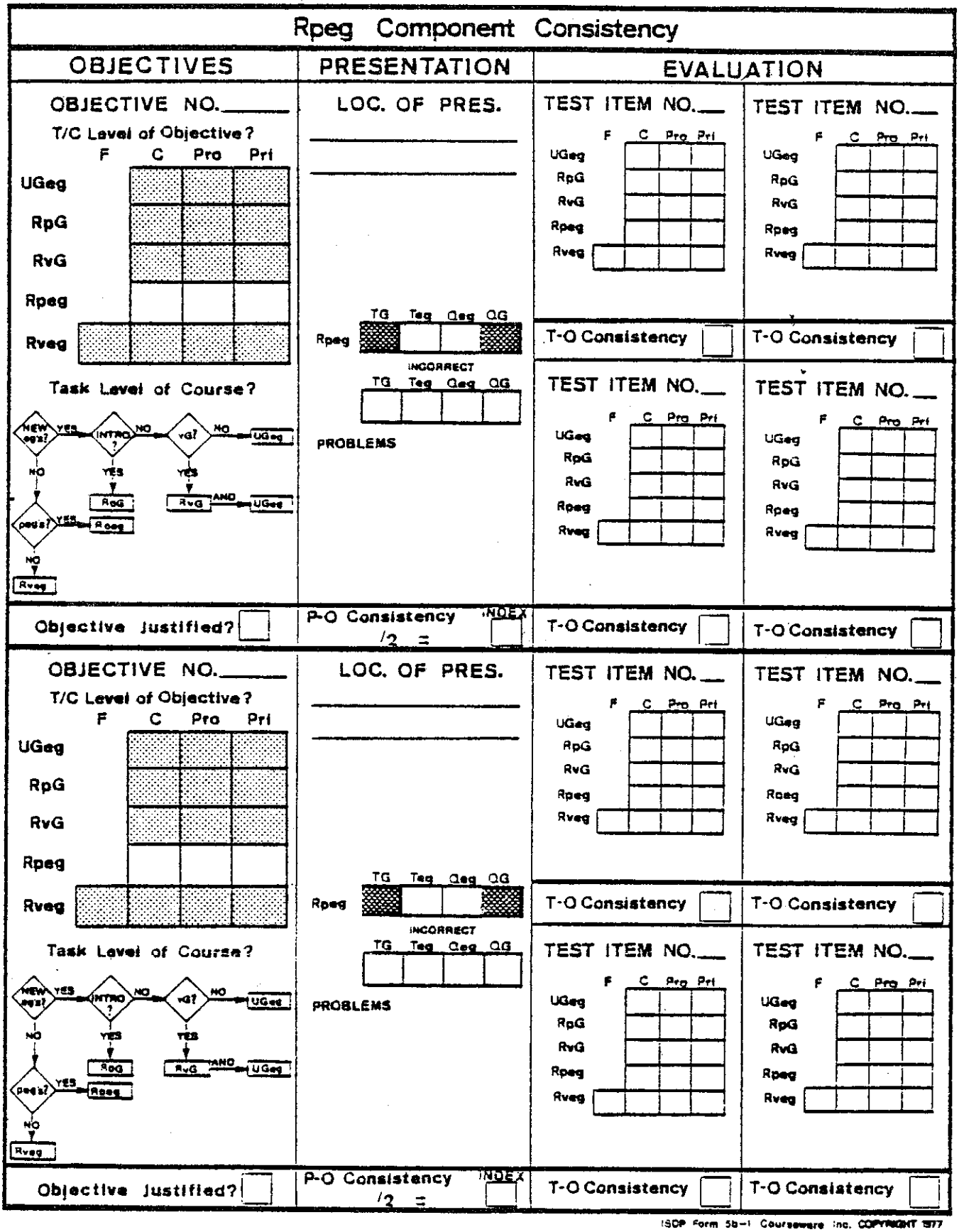


FIGURE 4. Sample ISDP form: Rpeg component consistency.

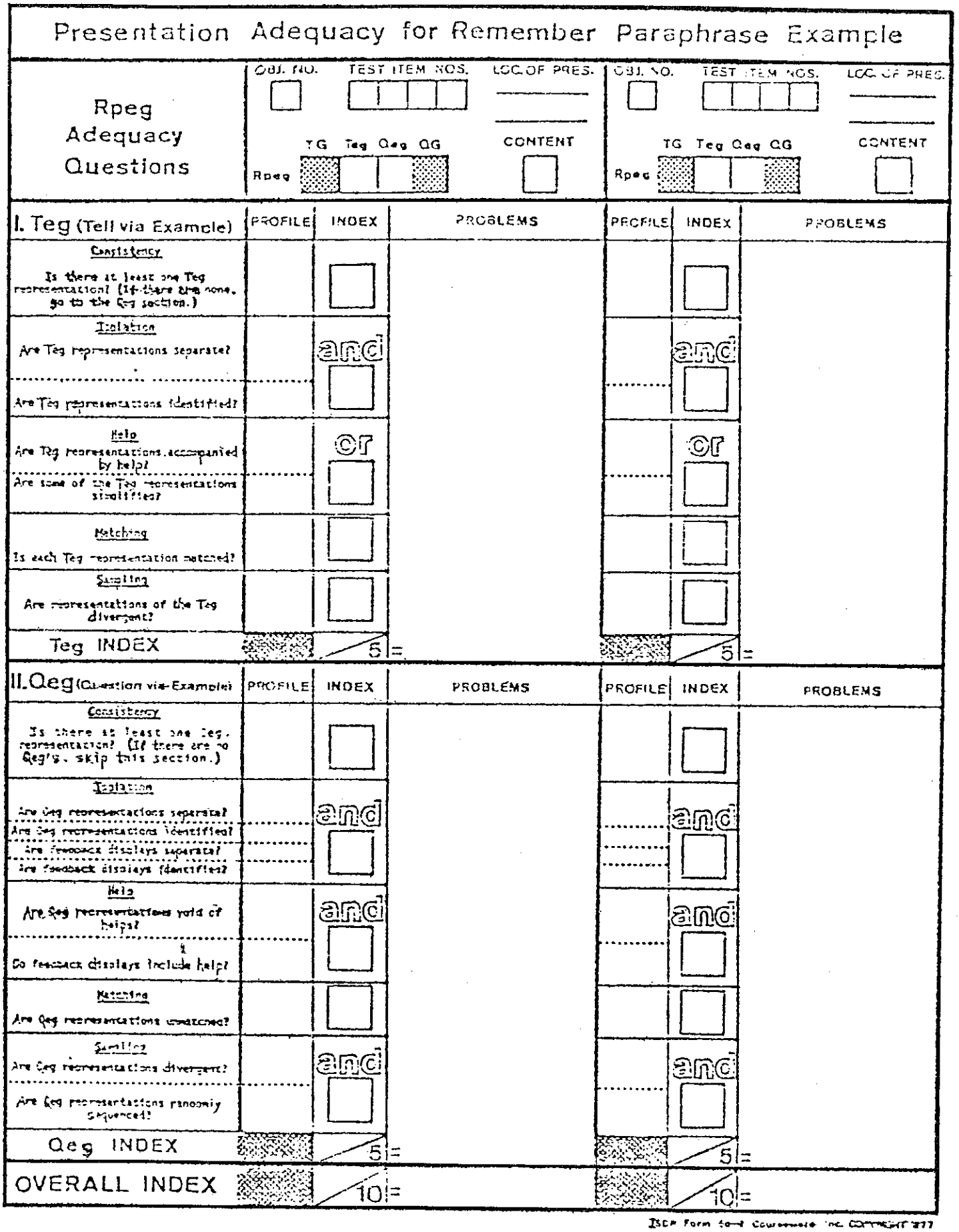


FIGURE 5. Sample ISDP form: Presentation adequacy for *remember paraphrase* example.

TABLE 1. ISDP ratings on the 18 selected test items

ACS test item	Morrison & Boyd text ISDP index				Streitwieser & Heathcock Text ISDP Index			
	TG	Teg	Qeg	OI	TG	Teg	Qeg	OI
2	1.00	.40	.00	.38	0	.20	0	.08
13	.67	.40	.00	.31	.33	0	0	.08
17	.33	.80	.80	.69	.33	0	0	.08
18[a]								
27	.67	.80	0	.46	.33	.40	.20	.31
29	.67	1.00	0	.54	.33	0	0	.08
33	.33	.80	0	.38	.67	.60	.40	.54
40	.67	0	.40	.31	0	.60	0	.23
43	1.00	.60	0	.46	.67	.60	.49	.54
48	.33	0	0	.08	.33	0	0	.08
54	.33	.60	0	.31	.67	.20	0	.23
56	.67	.60	.40	.54	.67	.20	0	.15
57	.33	.60	0	.31	.67	.40	0	.31
59	1.00	.60	0	.46	0	.40	0	.15
62	.67	0	0	.15	.33	.60	0	.31
66	.67	0	0	.15	0	0	0	0
68	.67	.80	.60	.69	.33	.40	0	.23
69	.33	.60	.20	.38	.33	.40	0	.23

Note. For the Morrison & Boyd text, $\bar{x}$ = .3882 and SD = .0432. For the Streitwieser & Heathcock text, $\bar{x}$ = .2135 and SD = .0388. TG = tell via generality. Teg = tell via example. Qeg = question via example. OI = overall index.

[a]Item number 18 has no ISDP ratings reported because no presentation could be found in either text on the content of the item.

TABLE 2. ACS organic chemistry test performance

	Morrison & Boyd Text	Streitwieser & Heathcock Text
	Winter 1978	Winter 1977
Teacher 1	$\bar{x}$ = 46.0 SD = 1.04 n = 77	$\bar{x}$ = 46.42 SD = 1.46 n = 52
	Winter 1978	Winter 1977
Teacher 2	$\bar{x}$ = 47.5 SD = 1.12 n = 74	$\bar{x}$ = 43.4 SD = 1.36 n = 49

Note. t = 1.374, $p < .10$. For the Morrison & Boyd text, mean = 46.66 and SD = .760. For the Streitwieser & Heathcock text, mean = 44.96 and SD = 1.000.

$p < .005$); the respective means are .3882 and .2135.[3] (See Table 1.)

Hypothesis 2 stated: "There is a significant difference on test performance between students using the two different texts." The results of this study show partial support for the above hypothesis. Students using the Morrison and Boyd text (1973) had higher test performance than the students using the Streitwieser and Heathcock text (1976) on the 70-item posttest (t = 1.3745, $p < .10$); the means respectively are 46.66 and 44.96. (See Table 2.) Also noted were marginal differences between the two groups on the 18 selected items (t = 1.52, $p < .10$); the means respectively are 61.25 and 55.08. (See Table 3.)

Hypothesis 3 stated: "There is a significant linear relationship between ISDP ratings and test performance of the two texts." The above hypothesis was not supported by the results of the study (t = -0.057, $p > .10$). The regression equation was y = 59.99 - 6.852x. See Figure 6 for a plot of the data.

Hypothesis 4 stated: "There is a significant difference between the text selected by the Chemistry Department and the text with the lower ISDP rating and lower performance." The Chemistry Department selected the Morrison and Boyd text. The results of this study tend to support the above hypothesis. Students given the Morrison and Boyd text outperformed overall the students given the Streitwieser and Heathcock text and the Morrison and Boyd text had higher ISDP ratings than the Streitwieser and Heathcock text.

Discussion

The Morrison and Boyd text (1973) received higher ISDP ratings than the Streitwieser and Heathcock text (1976), and although student performance was only marginally significant, it was in the expected direction. That is, the text with the higher ISDP ratings, the Morrison and Boyd text, also produced higher student performance. Other available data similar to the data of this study were analyzed. A separate analysis was conducted on this data because the authors felt that the subject populations were characteristically different. The results of this analysis fully support the second hypothesis (t = 2.137, $p < .025$), with the results in the expected direction. The Morrison and Boyd text group outperformed the Streitwieser and Heathcock group. The respective means were 48.26 and 43.44.

Although a linear relationship between posttest performance and ISDP rating was hypothesized, this particular research study found no such relationship. Merrill, in previous research (Note 2), has found that with similar content a linear relationship may not exist. In particular, the subject matter Merrill investigated contained test items of considerable complexity that the ISDP rating form possibly could not handle. The ISDP forms are designed to index one-to-one correspondence; however, certain subject areas do not follow a one-to-one correspondence between presentation and test items. Merrill has found a course in physics to be in this category (Note 3). Several of the items in the Organic Chemistry Exam were of

[3]The ISDP ratings are based upon 0.00 to 1.00 scale.

TABLE 3. ISDP ratings and test performance in organic chemistry course on the selected 18 test items

ACS test item	Morrison & Boyd text ISDP index[a]				Streitwieser & Heathcock text ISDP index[a]				Number of Correct Responses & Percent[b] Teacher 1				Teacher 2			
									Winter 1978		Winter 1977		Winter 1978		Winter 1977	
	TG	Teg	Qeg	OI	TG	Teg	Qeg	OI	#[c]	%	#[d]	%	#[e]	%	#[f]	%
2	1.00	.40	.00	.38	0	.20	0	.08	68	88	47	90	69	93	43	88
13	.67	.40	.00	.31	.33	0	0	.08	43	56	38	73	53	72	29	59
17	.33	.80	.80	.69	.33	0	0	.08	50	65	25	48	55	74	23	47
18									60	78	30	58	43	58	27	55
27	.67	.80	0	.46	.33	.40	.20	.31	30	39	17	33	24	32	15	32
29	.67	1.00	0	.54	.33	0	0	.08	40	52	17	33	36	49	21	43
33	.33	.80	0	.38	.67	.60	.40	.54	39	51	15	29	28	38	16	33
40	.67	0	.40	.31	0	.60	0	.23	59	77	38	73	65	88	40	82
43	1.00	.60	0	.46	.67	.60	.49	.54	46	60	31	60	59	80	27	55
48	.33	0	0	.08	.33	0	0	.08	54	70	28	54	48	65	27	55
54	.33	.60	0	.31	.67	.20	0	.23	37	48	28	54	48	65	10	20
56	.67	.60	.40	.54	.67	.20	0	.15	22	29	28	54	35	47	25	51
57	.33	.60	0	.31	.67	.40	0	.31	56	73	29	56	59	80	25	51
59	1.00	.60	0	.46	0	.40	0	.15	63	82	36	69	52	70	36	73
62	.67	0	0	.15	.33	.60	0	.31	50	65	36	69	36	49	30	61
66	.67	0	0	.15	0	0	0	0	24	31	27	52	27	36	20	41
68	.67	.80	.60	.69	.33	.40	0	.23	33	43	33	63	30	41	27	55
69	.33	.60	.20	.38	.33	.40	0	.23	64	83	35	67	68	78	23	47
			$\bar{x}$ = .3882				$\bar{x}$ = .2135		$\bar{x}$ = 60		$\bar{x}$ = 57		$\bar{x}$ = 62		$\bar{x}$ = 53	
			SD = .0432				SD = .0388		SD = 4.39		SD = 3.77		SD = 4.57		SD = 4.11	

Note. For the Morrison & Boyd text, $\bar{x}$ = 61.25. For the Streitwieser & Heathcock text, $\bar{x}$ = 55.08. TG = tell via generality. Teg = tell via example. Qeg = question via example. OI = overall index.

[a] $t = 3.104$, $p < .005$.

[b] $t = 1.52$, $p < .10$.

[c] Possible number of correct responses = 77.

[d] Possible number of correct responses = 52.

[e] Possible number of correct responses = 74.

[f] Possible number of correct responses = 49.

the same type as described by Merrill.

Apparently the approach taken in this study can be helpful in textbook selection. The text that had higher ISDP rating and higher posttest student performance was also the text selected by the Chemistry Department.

Recommendations

Because a linear relationship was not found with this data and because the authors feel that the ISDP form may be limited to one-to-one correspondence type materials, it is recommended that further analysis be conducted on those test items from this study which are one-to-one correspondence types. The expectation is that the analysis will find a significant linear relationship for those items which have a one-to-one correspondence. Such research would be helpful in predicting student performance based on one-to-one correspondence type materials ISDP ratings.

On the other hand, Wood, Richards, and Merrill (1976) have conducted research investigating complex objectives or test items similar to those found in this study using ISDP rating forms. Wood used a complexity of performance task measure defined as the number and type of task/content relations inherent in individual test items. (His findings indicate that a high correlation exists between student performance ratings and ISDP ratings of the presentation materials.) Therefore, we suggest that the present materials be reassessed using the complexity of performance task measure employed by Wood in his study. The expectation is that the more complex the task/content relations are the higher the probability of incorrect manipulation of one or more task/content. Such research would enable developers to isolate on those test items where problems arise to do the following: (a) Identify items that should be simplified or broken down to smaller components of evaluation, if possible, (b) conduct "rear-end" analysis for diagnostic purposes, and (c) provide information as to what needs to be taught and included in the instructional materials.

The authors recommend other instructional materials be likewise studied, especially when material selection decisions are to be made. Also, research on teacher effects (i.e., supplemental materials, specific emphasis, laboratory experiences, etc.), which were not controlled in this present study need to be addressed.

The authors suggest that all instructional materials be rated according to the ISDP and that areas of low ratings receive development through a joint contribution from content experts and ISDP experts.

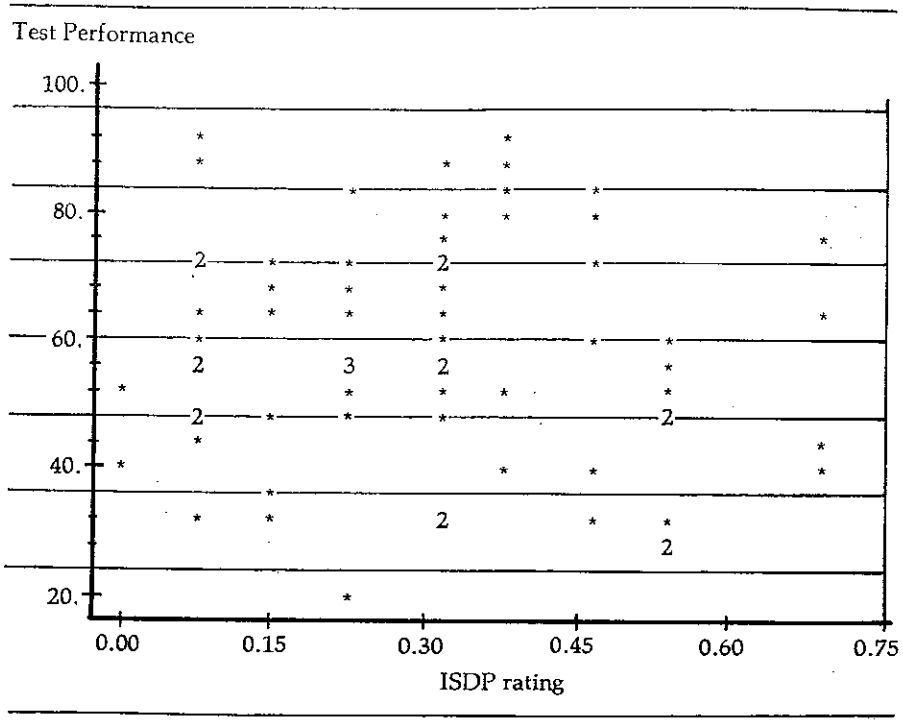


FIGURE 6. Plot of the correlation of the ISDP ratings and test performance.

Reference Notes

1. Merrill, M. D., Richards, R. E., Schmidt, R. V., & Wood, N. D. *The instructional strategy diagnostic profile training manual* (NPRDC Spec. Rep. 77-14). San Diego: Navy Personnel Research and Development Center, September 1977.
2. Merrill, M. D. Personal communication, April 1979.
3. Merrill, M. D., Schmidt, R. V., & Norton, R. F. *Validation of the instructional strategy diagnostic profile in physics 100* (Instructional Rep. No. 64). Provo, Utah: Courseware, Inc., 1977.

References

Merrill, M. D., Olsen, J. B., & Coldeway, N. A. Research support for the instructional strategy diagnostic profile. *Technical Report No. 3.* Provo, Utah: Courseware, Inc., 1976.

Morrison, R. T., & Boyd, R. N. *Organic Chemistry.* Boston: Allyn and Bacon, Inc., 1973.

Streitwieser, A., Jr., & Heathcock, C. H. *Introduction to Organic Chemistry.* Macmillan Publishing Co., Inc., 1976.

Wood, N. D., Richards, R. E., & Merrill, M.D. *Prediction of Student Performance on Rule Using Tasks from the Diagnosis of Instructional Strategies.* Provo, Utah: Brigham Young University Research Paper, 1976.