SURVEY OF TEST SELECTORS' CONCERNS AND THE TEST SELECTION PROCESS

Margeret Dotseth
Russell Hunter
Clinton B. Walker

CSE Report No. 107

March 1978

Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California 90024

The activity which is the subject of this report was supported in part or in whole by the National Institute of Education, Department of HEW. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

SURVEY OF TEST SELECTORS' CONCERNS AND THE TEST SELECTION PROCESS

Margeret Dotseth
Russell Hunter
Clinton B. Walker

This report presents the results of a survey conducted as part of the development of the CSE Criterion-Referenced Test Evaluations, a resource book containing evaluative and descriptive reviews of sixty-four criterion-referenced tests. Since the book is intended to fill the information needs of education involved in selecting tests, an aim of the survey was to find out the information needs of test selectors to guide the design of the book. Three aspects of test selection were targetted: (1) the factors that affect test purchasers' choices among available tests, (2) consumers' knowledge, beliefs and attitudes about CRTs, and (3) the test selection process.

The survey was intended not only to identify which factors are important to test selectors but also to identify which factors might, owing to their unfamiliarity or unpopularity, need more explanation if they were to be included in the book. A second aim of the survey was to identify aspects of the context in which test selection occurs, in order for the book to address that context. Two subsidiary purposes were to identify a market for the book by finding out who selects tests and to promote interest about the CRTEB in that market.

Development of the Survey

Preparation of the first draft of the instrument was guided by advice from staff members of three school districts. The Center's previous market

studies were also consulted. After a series of items dealing with test selection and test evaluation formats had been generated, they were reviewed by two members of the Los Angeles Unified School District (LAUSD) staff. Approval was obtained to distribute the instrument under a LAUSD cover letter to a sample of elementary school principals.

This pilot version of the survey was sent to 50 principals, of whom only 10 (20%) responded. The low response rate was probably due to the low involvement of these principals in test selection: seven of the ten respondents indicated that they take part in selecting less than 20% of the commercially published tests used in their schools.

In spite of their lack of involvement in the test selection process, the respondents did provide important information about prevailing attitudes and practices in the schools. Three trends were apparent in the responses: (1) all ten favored the use of comparative scores such as percentiles, over absolute scores such as percent correct, for reporting the results of achievement tests; (2) respondents also expressed a need for the type of information that can be provided by criterion-referenced tests (i.e., sensitivity to instruction, and description of skills measured); (3) when asked to choose a preferred format style for the Criterion-Referenced Test Evaluation Book, all of the nine persons responding to that item chose a format combining a detailed tabular layout with prose descriptive and evaluative summaries.

Several other inputs informed the development of the survey. The marketing department of CTB McGraw-Hill, Monterey, briefed project staff on their recent consumer survey. CTB had selected as respondents the same target

population which CSE had picked--namely, specialists in testing, curriculum, and guidance on school district staffs. Next, existing sets of standards for evaluating tests were reviewed, including the MEAN evaluation system (Hoepfner, 1976) and draft versions of Standards for Evaluating Criterion-Referenced Tests (Walker, 1978). These latter standards were in part a product of a review of the issues surrounding criterion-referenced measurement (CRM). Finally, the item pool was critically reviewed by an outside expert in CRM.

The final pool of items was formidable, consisting of the following: rating scales for 33 judgments in selecting tests; ratings of 11 sources of information on tests; 18 open-ended questions on the social process of test selection; and over 100 questions about educators' knowledge and beliefs of CRTs. In order to produce surveys of manageable length, two steps were taken. First, the items dealing with educators' knowledge and beliefs were saved for another CSE project on test use (Yeh, 1978). Second, the remaining item pool was divided into two survey instruments. One form, which will be referred to as the factors survey, dealt with test characteristics and sources of information which affect choices among test. Appendix I is a copy of that instrument. The other form, called the process survey, asked about the identities of test selectors and the process of selection. To determine how widely CRTs are used, a question was included on CRT use in the respondents' districts. The process survey is Appendix II.

Dissemination

Three thousand mailing labels were ordered from Curriculum Information Center, Denver. These consisted of random samples of 1000 educators from

each of the following occupational categories: testing; guidance and counseling; and curriculum and instruction. The two survey instruments were then sent to non-overlapping samples of 1500 educators, each form going to 500 people in a job category. The process survey was printed on a prestamped, mailable form, whereas the factors survey was accompanied by a prestamped return envelope.

Results

Appendices I and II present summary descriptive data for the 531 respondents returning the factors survey and 361 returning the process instrument by November 30, 1976. These figures represent return rates of 35% and 24%, respectively. After November 30, another 39 and 54 of the respective instruments were received, but those data are not included in the present analysis.

Factors Survey

In the 531 replies, the respondents' self-report of job or position was as follows:

12.2% testing or evaluation specialist

16.9% curriculum or instruction specialist

28.2% Guidance and Counseling specialist

11.3% principal or administrator

23% "Other"

7.7% two or more job categories

0.6% blank

The large number of "other" responses was due mainly to respondents being Directors, Supervisors, Coordinators or Assistant Superintendents in one

of the four categories rather than specialists, as asked.

Respondents were asked to rate the importance of various aspects of achievement tests on a five-point scale (1=barely, 3=moderately, 5=crucially). There was a tendency for respondents to give a relatively high importance rating to all features: only two of the 33 factors had mean ratings below 3.5 on the five-point scale. This consistency may reflect a positive response bias of some sort.

There was, however, noticeable differentiation of responses according to test characteristics and to respondents' job category. The least concern was shown for whether cultural bias had been minimized and for the availability of alternative forms, inservice training, and a record-keeping system (all means < 3.7). The greatest importance was given to the appropriateness of the test for intended examinees and to the nature of the skills being tested (means > 4.5). A somewhat lower concern (around 4.0) was shown for the technical aspects of tests, the relationship of the test to classroom management, and provision of a format for reporting scores to parents and students.

For each of the 33 factors a one-way analysis of variance of ratings was computed using job title or position as the independent variable. Fourteen of the 33 factors were rated significantly different by the various specialists. Table 1 presents a selection of data from this analysis.

Respondents were also asked to rate on a five-point scale their reliance on various sources of printed information in selecting tests. In descending order of importance the sources and their mean ratings were as follows:

specimen sets and technical manuals (4.32)

reference books that review tests (3.97)

lists of tests either reviewed or approved by the district or higher levels (3.29)

publishers' catalogs and advertising by mail (2.80)

publishers' ads in journals (2.09)

A portion of the respondents (21.5%) reported using other printed materials as well. Table 2 (items 34-39) gives a selection of data for the four printed sources whose rated importance varied significantly as a function of respondents' job category. Respondents were also asked to rate how much they relied on various personal contacts for information on tests, using the same five-point scale. They reported that they relied on the following groups, in descending order (mean ratings in parentheses):

school district staff (3.88)

classroom teachers (3.60)

principals (3.48)

publishers' representatives (3.05)

About one third of the respondents (32.4%) reported relying also on other categories of personal contacts for information on tests. Table 2 (items 40-43) gives a selection of data for the four types of personal resources of information whose rated importance in test selection varied significantly as a function of respondents' job category.

Process Questionnaire

Of the 361 surveys, the occupational breakdown is as follows:

145 (40%) Curriculum and Instruction

85 (24%) Guidance and Counseling

131 (36%) Testing

TABLE 1

Significant Differences in Rated Importance of Test Characteristics As a Function of Respondents' Job Category

Item Number and Characteristic	$p <$	*Lowest Rating	*Highest Rating
2. Bias	.025	G/C 3.5	C/I 3.9
6. Percentiles	.01	Other 4.0	G/C 4.4
9. Items keyed to Objectives	.05	G/C 4.1	C/I 4.5
10. Enough items per Objective	.01	G/C 3.9	C/I 4.3
13. Alternate forms	.01	G/C 3.1	C/I 3.6
15. Instructions to students	.01	G/C 4.1	P/A 4.6
17. Organization of technical manual	.01	G/C 3.7	P/A 4.3
18. Availability of in-service	.01	G/C 2.9	C/I 3.4
19. Sensitivity to instruction	.01	G/C 3.9	P/A 4.5
20. Curriculum cross-referencing	.01	G/C 3.6	P/A, C/I 4.1
29. Ease of responding	.025	Other 4.1	P/A, C/I 4.5
31. Classroom records system	.01	G/C 3.4	P/A 4.2
32. Parent report forms	.01	T/E 3.9	P/A 4.4
33. Group report forms	.01	T/E 3.5	C/I 4.0

*Abbreviations stand for the following job categories: T/E - Testing or Evaluation; C/I - Curriculum or Instruction; G/C - Guidance and Counseling; P/A - Principal or School Administrator; Other (no Abbreviation). On the 5-point scale, 1=barely important, 3=moderately important, and 5=crucially important.

TABLE 2

Significant Differences In Rated Reliance on Printed and Personal Sources of Information About Tests as a Function of Respondents' Job Category

Item Number and Information Source	$p <$	Value and Category of: *Lowest Rating	*Highest Rating
34. Publishers' Catalogs	.05	C/I 2.6	T/E 3.0
37. Reference books	.025	P/A 3.8	T/E 4.3
38. Test listings	.01	T/E 3.0	C/I 3.6
39. Other Printed	.01	P/A 2.1	T/E 3.2
40. Teachers	.05	Other 3.4	C/I 3.8
41. Principals	.01	T/E 3.3	P/A 3.9
42. Publishers' representatives	.01	G/C 2.8	P/A 3.5
43. District Staff	.01	G/C 3.6	C/I 4.2

*Abbreviations stand for the following job categories: T/E - Testing or Evaluation; C/I - Curriculum or Instruction; G/C - Guidance and Counseling; P/A - Principal or School Administrator: Other (no Abbreviation). On the 5-point scale, 1=not at all, 3=moderately, 5=heavily.

The respondents reported being involved in test selection for an average of about 9 years (median = 8.8, mean = 9.5).

The respondents were asked for which grade levels they selected tests. Ten percent reported K-6 responsibilities, 10% K-9, 13% 7-12, 55% K-12, with 12% reporting other grade level combinations (e.g., 4-9).

Respondents were asked to estimate the percentage of tests chosen by each of the following six categories of personnel: individual teachers acting alone; teachers within a school acting together; teachers with administrators in their own school; building level administrators working alone; district-level or district-wide personnel; and others. They reported (by mean percentage) that district personnel made the largest number of decisions (41%), followed by "other" (33.8%) personnel and in descending order, teachers acting with administrators in a school (26.5%), building level administrators working alone (16.2%), teachers within a school acting together (10.3%), and individual teachers alone (5.8%). The existence of test selection committees was reported by 62.3% of the respondents.

In selecting tests for their schools, 62% reported trying to match skills taught in their schools with skills covered by tests. Those reporting such an attempt were asked to describe how such matching was done. Fifty-five respondents reported that they matched the test to the curriculum, without much elaboration; 46 reported that matching was done "by committee"; 35 reported that test items, objectives or specimen sets were examined, and 17 reported that "item analysis" was done. A small group of respondents (5) noted the difficulty of the process by responses ranging from "it is difficult and not well done" to "seat of pants." Half of the respondents reported that they

or their co-workers often prefer to construct their own tests. In elaborating that response, twenty-three gave a time-linked answer (e.g., "during school year"). An equal number reported that tests were constructed as a need arose, 21 reported that tests were constructed for regular classroom use, and 10 reported that tests were constructed to match the curriculum. Thirteen respondents reported that while they currently did not construct their own tests, the possibility was under discussion. Item #11 asked if the educators had ever used a test that they would hesitate to use again, and, if so, to name the test and the problem. Half responded positively. The most common reasons cited were that the test was too lengthy (n=15) or that the test was unrelated to their curriculum (n=15). Other reasons included (in descending order of frequency): inappropriate norms, outdated material, test "too easy," tests inappropriate for their students, leveling inappropriate (this included comments on physical characteristics of the test), cumbersomeness, and poor validity. This item may have been too pushy for some respondents. A much larger number of respondents (36) skipped it than skipped the two other fill-in items that were discussed in the previous two paragraphs (7 & 8 respectively). Several responses indicated that the information was sensitive or confidential.

Local use of criterion-referenced tests was reported by 40% of respondents. In reporting highest level of detail in which curriculum is stated, 30% reported skill categories (e.g., arithmetic - basic operations), 28% general instructional objectives, 13% behavioral objectives, 2.2% amplified objectives and 2.5% other (note: 8% checked both skill categories and general instructional objectives). Eighty-seven percent of the respondents reported

consulting test reviews during the test selection process. Respondents indicating that they consult reviews were asked to indicate which of the following resources they consult: Buros, CSE, journals or "Other." Two-thirds of those responding indicated that they used two or more of the sources. Buros was reported as a source by 82.6% of those responding to this item, Journals by 57.2%, CSE Test Evaluations by 30.6%, and "other" sources by 15.1%. The CSE Test Evaluation Books were used largely in conjunction with other sources; only 1.6% of the respondents reported using the CSE test evaluations alone, 22.2% reported using only Buros, 7.3% journals alone, and 2.2% reported using "other" sources exclusively. (Note: the "other" category referred mainly to contacts in other schools or districts and publishers' materials.)

When asked how likely they would be to look up separate reviews of a test cited in a book of test reviews, 38% of the respondents reported that they would look up the review over half the time, 28% reported less than half the time, and 27% reported half the time (7% did not respond).

Discussion

In the factors instrument (Appendix I) all test characteristics were rated to be at least moderately important. This result supported two decisions about the CRTEB. The first was to include considerable descriptive information about CRTs. Previous CSE Test Evaluations had given almost exclusively evaluative information. Second, we decided to give a justification for each of the standards for evaluating CRTs (Walker, 1978) in the book, and to give a justification for using CRTs which, at present, generally lack some of the characteristics which the respondents valued highly.

Over 40% of the test characteristics have an importance that varies significantly with the job category of the respondent. This fact is due in part to differences in the types of test which the job specialties are most interested in. The survey asked about achievement tests, since almost all CRTs are tests of maximum performance and are supposed to be sensitive to instruction. Out of 14 significant differences in ratings in Table 1, guidance and counseling specialists gave the lowest rating for 10 of them. Those specialists are probably more concerned with measures of aptitude, interest, and in some cases, personality, than are the other specialists.

In the process survey (Appendix II), the results for items 6 and 7 indicated that the CRTEB should stress the match between tests' objectives and the objectives of the local curriculum, and that the book should provide advice on how to maximize that match.

The two instruments provided valuable marketing information by identifying individual respondents who are interested in the CRTEB (n=700), categories of educators who are involved in test selection, and types of information which educators are looking for.

References

Hoepfner, R., et al. CSE Elementary School Test Evaluations. Los Angeles: Center for the Study of Evaluation, University of California, 1976.

Walker, C. B. Standards for Evaluating Criterion-Referenced Tests. CSE Report, Center for the Study of Evaluation Technical Report No. 103. Los Angeles: Center for the Study of Evaluation, University of California, 1978.

Yeh, J. P. Test use in schools. CSE Measurement and Methodology Program. Work Unit 4. Deliverable, June, 1978 to the National Institute of Education, Department of Health, Education and Welfare.

APPENDICES

Appendix I

TEST SELECTORS' FACTORS SURVEY
[with Summary data included]

A. How important is each of the following factors to you in choosing an achievement (both criterion-referenced and norm-referenced) test? (Circle a number for each.)

Feature	How important it is to you					
	barely		moderately		crucially	Mean
1. which ages or grade levels the test is intended for	1	2	3	4	(5)*	4.616
2. whether cultural bias has been minimized	1	2	(3)	4	5	3.678
3. how long each test or subtest takes	1	2	3	(4)	5	3.815
4. whether the test had field tryouts	1	2	3	4	(5)	4.449
5. whether grade level equivalents of raw scores are provided	1	2	3	(4)	5	3.927
6. whether percentile norms are provided	1	2	3	4	(5)	4.198
7. which educational objectives are covered by the test	1	2	3	4	(5)	4.592
8. how precisely the objectives or skills are described	1	2	3	4	(5)	4.343
9. whether the test items are explicitly keyed to specific skills or objectives	1	2	3	4	(5)	4.306
10. whether there are enough items per skill or objectives	1	2	3	(4)	5	4.164
11. whether the students in the field test and norm sample are a national cross section	1	2	3	4	(5)	4.177
12. whether the students in the field test and norm sample are comparable to your students	1	2	3	4	(5)	4.306

*note: circles indicate modal response on scales.

Feature	How important it is to you					
	barely		moderately		crucially	Mean
13. whether the test has alternate forms	1	2	(3)	4	5	3.374
14. how clear the instructions to the examiner are	1	2	3	4	(5)	4.304
15. how clear the instructions to the students are	1	2	3	4	(5)	4.697
16. whether layout, print, and illustrations are clear and easy to understand	1	2	3	4	(5)	4.553
17. whether the technical manual is well organized	1	2	3	(4)	5	3.951
18. whether inservice training for using the test is offered	1	2	(3)	4	5	3.123
19. how readily the test results show the effects of instruction	1	2	3	(4)	5	4.152
20. whether test items are indexed to curricular materials for prescriptive purposes	1	2	3	(4)	5	3.811
21. whether divergent validity (factor purity) is adequate	1	2	3	(4)	5	3.710
22. whether predictive validity is adequate	1	2	3	(4)	5	3.950
23. whether test-retest reliability is adequate	1	2	3	(4)	5	4.096
24. whether alternate form reliability is adequate	1	2	3	(4)	5	3.844
25. how representative the items are of the skill to be tested	1	2	3	4	(5)	4.489
26. how well the test spreads out students' scores	1	2	3	(4)	5	3.603
27. how accurately a test establishes a student's mastery or non-mastery	1	2	3	4	(5)	4.353
28. whether there are flaws in item construction (like specific determiners, lack of a correct answer, etc.)	1	2	3	4	(5)	4.325
29. how easy it is for students to record their answers	1	2	3	4	(5)	4.283

Feature	How important it is to you					
	barely		moderately		crucially	Mean
30. how easy and objective the scoring is	1	2	3	(4)	5	4.146
31. whether a record keeping system is provided for classroom level use	1	2	3	(4)	5	3.658
32. whether a good format is provided for reporting test results to students and their own parents	1	2	3	(4)	5	4.075
33. whether a good format is provided for reporting test results to groups of parents and to the community	1	2	3	(4)	5	3.711

B. When you select tests, how much do you rely on each of the following sources of information? (Circle a number.)

PRINTED INFORMATION	not at all		moderately		heavily	Mean
34. Publishers' ads in journals	1	(2)	3	4	5	2.138
35. Publishers' catalogs and advertising by mail	1	2	(3)	4	5	2.804
36. Specimen sets and technical manuals	1	2	3	4	(5)	4.316
37. Reference books that review tests (like Buros or CSE Test Evaluations)	1	2	3	(4)	5	3.968
38. Lists of tests either reviewed or approved by the district or higher levels	1	2	3	(4)	5	3.287
39. Other printed information (please specify):	(1)	2	3	4	5	2.627*

*note: 68.7% of respondents did not answer this item

PERSONAL CONTACTS

	not at all		moderately		heavily	Mean
40. Classroom teachers	1	2	3	(4)	5	3.596
41. Principals	1	2	3	(4)	5	3.485
42. Publishers' representatives	1	2	(3)	4	5	3.057
43. School district staff	1	2	3	(4)	5	3.877
44. Other people (please specify)	1	2	3	4	(5)	3.738*

*note: 63.3% of respondents did not answer this item

C. Please list any other factors that influence you to choose a test or to avoid choosing one: Factors specified by 61% of respondents

_____

_____

D. What is your position or job title? (Check one)

12.2% Testing or Evaluation Specialist

16.9% Curriculum or Instruction Specialist

28.2% Guidance and Counseling Specialist

11.3% Principal or School Administrator

23.0% Other (Please specify): _____

Note: 7.7% listed more than one title, 0.6% omitted item.

E. Thank you very much for sharing your experience with us. If you would like to receive a list of the names and developers of criterion-referenced tests, or if you would like to receive information on a book (due in mid-1977) that systematically lists, describes, and evaluates all available criterion-referenced tests, check the respective blanks below and your address.

74.6% Please send me the list of CRTs.

77.4% Please send me information on the CRT Evaluation Book, when available.

Name and Title: _____

Address: _____

_____ Zip _____

Appendix II

TEST SELECTORS' PROCESS SURVEY
[with summary data included]

1. How long have you been involved in test selection?

   9.5 years. [mean] median=8.8

2. For which levels do you take part in choosing published tests? (Check all that apply.)

   ____ K-3 ____ 4-6 ____ 7-9 ____ 10-12

3. How large is your school district?

   9,988 students [mean] median=3,950

4. Out of all the commercially published tests used in your district, estimate the percentage that are chosen by each of the following:

	mean		median
-individual teachers acting alone	5.8	%	2.4
-teachers within a school acting together	10.3	%	4.6
-teachers with administrators in their own school	26.5	%	17.0
-building level administrators working alone	16.2	%	5.3
-district level or district-wide personnel	41.0	%	53.7
-other _____ (please specify)	33.8	%	17.2
Total	100	%	

5. If your district has a test selection committee, what positions do the members hold?

   Answered by 225 (62.5%) of respondents

6. How detailed is the curriculum for your schools? (Check the most detailed form in which your curriculum is stated)

   30%** skill categories (for example, in arithmetic: basic operations, measurement, problem solving, geometry, etc.)

   28%** general instructional objectives (for example: students will be able to comprehend assigned reading material; students will be able to tell time)

   13% behavioral objectives

   2.2% amplified objectives (or domain specifications or item forms)

   2.5% other (please specify) _____

   **8% of respondents checked both categories

7. In selecting tests for your schools, is an effort made to match the skills covered by the available tests with the skills that are actually taught? (Check one)

62% Yes 36% No (If NO, skip to #9)

8. How is this matching done?

see text

9. Do you and your co-workers ever prefer to construct your own test rather than use a published test? (Check one)

50% Yes 46% No (If NO, skip to #11)

10. When do you construct your own? see text

11. Have you ever used a particular published test that you would hesitate to use again?

50% Yes 39% No (If NO, skip to #13)

If so, what was the test and what was the problem? see text

13. In the process of selecting tests do you ever consult reviews of tests?

87% Yes 10% No (If No, skip to #16)

14. In which sources? (Check all that you use)[N. B. Reported percentages are based on proportion of responses to this item.]

83% Buros 31% CSE Test Evaluations 57% Journals

15% Other see text
please name

15. If a book of test reviews made reference to other reviews (for example, in journals) of the same tests, would you be likely to look up the other reviews? (Circle one)

Mean = 3.17

1	2	3	4	5
rarely		half the time		often
[12%]	[16%]	[27%]	[20%]	[18%]

Missing [7%]

16. Are criterion-referenced tests used in your district?

40% Yes 51% No (If NO, skip to #18)

17. Which one(s)?

name	source or publisher

Answered by 141 (39.1%) of respondents

18. If you would like to add any comments on how tests are selected, practical alternatives to testing, or anything else, please put them here.

Comments made by 84 (23.3%) of respondents.

Summary of respondents by job category:

Job Category	Number	Percentage	
Curriculum & Instruction			
K-12	69	19%	40%
K-6	48	13%	
7-9	28	8%	
Guidance & Counseling			
K-12	61	17%	
K-6	6	2%	24%
7-9	18	5%	
Testing			
All grades	131	36%	