Achievement Test Selection: Making the Best Choice

Lore A. Nielsen, John A. Hilderbrand, Susan D. Turner
Hillsborough County Public Schools
Tampa, Florida

Selecting Your District's Achievement Test: How to Make the Best Choice

That system-wide normative testing is an educational fact of life is undeniable; the demands of accountability ensure its continued well-being. Furthermore, where accountability has been demanded, litigation has quickly followed. Districts have been required to demonstrate in court that tests measure what has been taught in the classrooms. Given these realities, however, how do decision-makers at the district level determine which test among the many touted by publishers best meets their needs?

One large Florida school district, Hillsborough County, followed a set of procedures in their test selection process which could be used by other districts facing a similar dilemma. These procedures were established by the Department of Testing and Evaluation and sanctioned by the Directors of Instruction. In formulating the procedures, the first consideration was the purposes for which the selected test would be used. By looking at the way the current test was used, Testing and Evaluation personnel determined that results would form the basis for decisions on classroom instruction, school-wide curriculum modifications, and county-level planning. The overriding purpose of testing at all these levels was clearly improvement of instruction. Therefore, throughout the test selection process, emphasis was placed on finding the test which most closely matched Hillsborough County's curriculum.

The current norm-referenced test had been used since 1975 and scores had increased steadily. The normative data from the test were ten years old and no longer provided a sound comparison of district students with a national sample. Moreover, high student and teacher familiarity with test items might be causing a test-retest phenomenon. The most critical problem, however, was that no evidence existed to support the premise that the norm-referenced test reflected the county's present curriculum. Although a match between the two may have existed when the test was adopted, there was no reason to assume the test continued to assess achievement related to the district's curriculum.

The review process for selecting a new standardized test was a multi-step process. Activities fell into three main categories:

1. Preparation for test review,
2. Test presentations, and
3. Data analysis and field testing.

Preparation for Test Review

The six largest test publishers were given and accepted an invitation to participate in the review process. They were then sent materials describing the school system including the most recent achievement report, socio-economic data, and curriculum objectives. They were also provided with the review form which would be used and guidelines in choosing the test they would present: 1) the test must match the county curriculum, grades K through

10; 2) the test had been developed using acceptable methodology; and 3) it had been normed no earlier than 1977.

Each test publisher was assigned a presentation date and provided guidelines for the presentation which included length of presentation, place, and committee composition. Publishers were encouraged to emphasize the curriculum-test match and available reporting procedures. Each publisher was requested to provide to the Department of Testing and Evaluation two weeks prior to the presentation sufficient specimen materials for each member to receive the appropriate level tests and related materials.

Moreover, publishers were cautioned to make all contacts concerning the test review through the Department of Testing and Evaluation; that is, no instructional persons were to be contacted privately and/or individually about the test selection process.

While communication with the six publishers was ongoing, the Director of Testing and Evaluation formed a steering committee of four people he felt could give the most input in the planning process. None of these persons served on the subsequently appointed test review committee. This group reviewed current literature on test selection procedures as well as test review forms used by other school systems and made recommendations for the review procedures. The Director of Testing and Evaluation and his staff drafted the forms based on their suggestions. Five areas were covered by the review forms: test content, quality of items, test format, test administration and reports. (See Appendix A for a copy of the review form.)

Under test content, determination was made of how well the test matched the district's curriculum and State Minimum Standards. How thoroughly the cognitive domain had been sampled, how closely items matched the curriculum, and the extent of curriculum coverage on the test were issues to be decided. Examination of the quality of items included considering the appropriateness of the test in assessing the students for whom it was intended, including readability of items, control of bias, relevance of items to students' experiences, item difficulty, and item clarity.

Test format, or the physical make-up of the test, included consideration of use of illustrations, print size, items per page, item arrangement, use of color, clarity of test directions within the test, and sample items. The area of test administration included time allotment, directions for test administration, and the informational content of the manual. Finally, the review of the reports possible with each test entailed evaluating the completeness of information contained in the reports, curriculum decisions which would be made on the basis of each test, availability of subtest analysis, clarity of reporting forms, and facilitation of interpretation of test data.

Selection of a test review committee was the next logical step in the process. The committee was designed so that all grades, levels of instructional decision-makers, and impacted subject areas were represented fairly. Positions represented on the committee were: principals, counselors, assistnat principals, curriculum specialists, supervisors and teachers. The Testing and Evaluation Director selected persons to fill the committee positions randomly from lists of individuals recommended by either the appropriate Director (or General Director) or the principals' council. Each person was given responsibility for reviewing certain levels of the test in one or both of two subject areas: mathematics or communications (see Figure 1 for test review assignments). Personnel in the Department of Testing and

Figure 1

Test Review Assignments

Grade:	K		1		2		3		4		5		6		7		8		9		10	
Subject / Position	M	C	M	C	M	C	M	C	M	C	M	C	M	C	M	C	M	C	M	C	M	C
Principal (Elem.)	x	x	x	x	x	x	x	x	x	x	x	x	x	x								
Principal (Jr. High)															x	x	x	x	x	x		
Principal (Sr. High)																					x	x
Asst. Principal (9-12)																						
Supervisor (Early Child)	x	x	x	x	x	x	x	x														
Supervisor (Elem. Math)	x		x		x		x		x		x		x									
Supervisor (Elem. Read.)		x		x		x		x		x		x		x								
Supervisor (Sec. Math)															x		x		x		x	
Supervisor (Sec. Comm.)																x		x		x		x
Supervisor (DEES)	x	x	x	x	x	x	x	x	x	x	x	x	x	x								
Counselor (Elem.)	x	x	x	x	x	x	x	x	x	x	x	x	x	x								
Counselor (Jr. High)															x	x	x	x	x	x		
Counselor (Sr. High)																					x	x
Curriculum Specialist	x	x	x	x	x	x	x	x	x	x	x	x	x	x								
Curriculum Specialist																	x	x	x	x	x	x
Teacher (Primary)	x	x	x	x	x	x	x	x														
Teacher (Intermediate)									x		x		x									
Teacher (Intermediate)										x		x		x								
Teacher (Jr. High Math)															x		x		x			
Teacher (Jr. High Comm.)																x		x		x		
Teacher (Jr. High Comm.)																x		x		x		
Teacher (Sr. High Math)																					x	
Teacher (Sr. High Comm.)																						x
TOTAL	7	7	7	7	7	7	7	7	6	6	6	6	6	6	4	5	5	6	5	6	5	5

M = Mathematics
C = Communications

53

the cumulative mean across the areas were used to compare strength and weaknesses of the two tests. These ratings were also used to give feedback to the test publishers about the committee's reactions to their tests.

The Director of Testing and Evaluation and his staff next reviewed the two tests identified by the test review committee for technical qualities. This review involved examining the technical manual for data on test construction procedures, reliability, validity, norming procedures, etc., as well as analyzing the theoretical models on which the tests were based. The decision had been made in planning the entire selection process to leave this review until after the tests which best matched the curriculum were identified. This decision was based on the assumption that all six of the largest publishers followed basically sound technology in constructing their tests. The technical review confirmed that assumption for the two tests analyzed.

At this point, the district's data processing personnel were consulted to determine the time and cost involved to establish the scoring and reporting packages. Negotiations over costs and services were initiated with each publisher.

One of the tests was then pilot tested with several randomly selected groups of students at grades three, five, and eight. Because normative data for the second test were not yet available, it could not be scored meaningfully and therefore was not piloted. For the test which was piloted, grade equivalent scores were compared to the most recent administration of the current test by grade level on similar subtests. These data along with all other pertinent data were considered in the Director's final recommendations.

Discussion

The Director of Testing and Evaluation made a final recommendation to the Assistant Superintendent for Instruction. The Assistant Superintendent's decision was later approved by the Superintendent and the Board. The chosen test will be used for the first time in spring 1983.

The Hillsborough County Department of Testing and Evaluation learned much during the test selection process. Some of the procedures followed worked well; some will be modified when it becomes time again to select a standardized test. First, involving an exceptional student supervisor was fortuitous; this person provided invaluable insights into color and format considerations based on knowledge of SLD, EH, and EMH students. In contrast, one group which should have been represented but was not was the Federal Programs division.

Separating the reviewers into elementary and secondary groups without regular articulation between the two areas caused some problems at the final meeting; each group preferred different tests. Had communication between them been ongoing, perhaps each group could have understood the position of the other and a test satisfactory both to elementary and secondary concerns could have been identified more readily.

Finally, the inservice afforded the reviewers was a serendipitous effect indeed. Supervisors and teachers alike gained knowledge about test construction and test considerations which they were able to add to their store of professional knowledge.

The test selection procedures followed by this district were long and involved, but they did permit an important decision to be made on the basis of sound data rather than on whim. Throughout the process, instructional considerations were emphasized over financial concerns. The result should be a testing program which matches the county's curriculum and yields data from which solid educational decisions can be made.

STANDARDIZED TEST REVIEW FORM
HILLSBOROUGH COUNTY SCHOOLS
1981

General Directions:

1. All information requested at the top of each form should be completed before starting the review process.

2. **Subtests:** This column contains the names of subtests normally found in standardized achievement tests. The subtests are grouped by major areas: Total Reading (1), Total Language (2), and Total Mathematics (3). If other or additional subtests exist, please list each at the bottom in the area labeled "Other". Assign the area 1, 2, or 3 and write the title of the subtest.

3. **Comments:** This column deals with all information you feel is important and has not been asked for already.

Test Content:

Test content should be reviewed by how well it matches the Hillsborough County curriculum and State Minimum Standards. Each test should be reviewed in relation to the grade(s) the test purports to assess.

A. **Fill in if Different Name:** This column is to be filled out only if the name of the subtest is different from the name given in the first column but the subtest meaning has not changed. (i.e., "Word Knowledge" could be listed as "Vocabulary.")

B. **Sampling of Cognitive Domain:** This column deals with the levels at which items measure the concepts covered in the test. For example, are all items asking the student to recall facts or are the students asked to apply or synthesize information? (Refer to Bloom's Taxonomy of the Cognitive Domain.) For a good rating in this column, higher as well as lower levels should be represented.

C. **Item-Match With Curriculum:** This column deals with the extent to which the items are consistent with the curriculum. Is the content of the item normally taught in that grade in Hillsborough County Schools?

D. **Extent of Curriculum Coverage:** This column deals with whether the items represent the range of the curriculum continuum within the grades the test purports to measure. This is different than the item-match in that every item could be included in the curriculum and yet not cover the breadth of the curriculum continuum.

Rater's Name _____	Test Name _____	Portions Reviewed	Key
Position _____	Test Level _____	Reading _____	0 - Poor/nonexistent/ina...
	Grades Covered _____	Language _____	1 - Weak/somewhat inapp...
		Mathematics _____	2 - Adequate/appropriate
			3 - Good/very appropriate

Test Content: Subtests	(A) Fill in if Diff. Name	(B) Sampling of Cognitive Domain	(C) Item-Match W/Curr.	(D) Extent Curr. Coverage	Comments
Word Knowledge	_____	0 1 2 3	0 1 2 3	0 1 2 3	
Reading Comp.	_____	0 1 2 3	0 1 2 3	0 1 2 3	
(1) Total Reading		0 1 2 3	0 1 2 3	0 1 2 3	
Spelling	_____	0 1 2 3	0 1 2 3	0 1 2 3	
Language Arts	_____	0 1 2 3	0 1 2 3	0 1 2 3	
(2) Total Language		0 1 2 3	0 1 2 3	0 1 2 3	
Computation	_____	0 1 2 3	0 1 2 3	0 1 2 3	
Concepts	_____	0 1 2 3	0 1 2 3	0 1 2 3	
Problem Solving	_____	0 1 2 3	0 1 2 3	0 1 2 3	
(3) Total Mathematics		0 1 2 3	0 1 2 3	0 1 2 3	
Others					
( ) _____	_____	0 1 2 3	0 1 2 3	0 1 2 3	
( ) _____	_____	0 1 2 3	0 1 2 3	0 1 2 3	
_____	_____	0 1 2 3	0 1 2 3	0 1 2 3	

TEST REVIEW FORM (Pg. 2)

General Directions:

1. All information requested at the top of each form should be completed before starting the review process.

2. **Subtests:** This column contains the names of subtests normally found in standardized achievement tests. The subtests are grouped by major areas: Total Reading (1); Total Language (2), and Total Mathematics (3). If other or additional subtests exist, please list each at the bottom in the area listed "Other". Assign the area 1, 2, or 3 and write the title of the subtest.

3. **Comments:** This column deals with all information you feel is important and has not been asked for already.

Quality of Items:

Quality of items should be reviewed in terms of the appropriateness for students within the grade(s) the test purports to assess.

A. **Readability of Items for Grade(s) Covered:** This column deals with whether the items are written at a level students will understand. Are the words used and sentence construction too difficult, too easy for the majority of the students?

B. **Control of Bias:** This column deals with whether the items frequently stereotype sex roles or ethnic groups, use words unique to specific regions of the country, etc. Such items would indicate bias, and the test would therefore be rated with a low number.

C. **Item Relevance to Student Experiences:** This column deals with whether the items are written in terms of events or situations which are likely to be familiar to the student.

D. **Item Difficulty:** This column deals with whether the items are written at a level which is too difficult or too easy in terms of the material used or the concepts represented. A good test will have both hard and easy items with the majority of items being of average difficulty.

E. **Item Clarity:** This column deals with whether the student understands what the item is asking the student to do. In addition, each multiple choice item should contain distractors which are plausible answers.

Test _____

Level _____

Rater _____

Key

0 - Poor/nonexistent/inappropriate

1 - Weak/somewhat inappropriate

2 - Adequate/appropriate

3 - Good/very appropriate

Quality of Items: Subtests	(A) Readability for Grade(s) Covered	(B) Control of Bias	(C) Item Relevance to Student Experiences	(D) Item Difficulty	(E) Item Clarity	Comments
Word Knowledge	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Reading Comp.	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
(1)Total Reading	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Spelling	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Language Arts	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
2)Total Language	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Computation	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Concepts	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Problem Solving	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
3)Total Mathematics	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Others						
( ) _____	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
( ) _____	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
( ) _____	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	

TEST REVIEW FORM (Pg. 3)

__General Directions:__

1. All information requested at the top of each form should be completed before starting the review process.

2. __Subtests:__ This column contains the names of subtests normally found in standardized achievement tests. The subtests are grouped by major areas: Total Reading (1), Total Language (2), and Total Mathematics (3). If other or additional subtests exist, please list each at the bottom in the area labeled "Other". Assign the area 1, 2, or 3 and write the title of the subtest.

3. __Comments:__ This column deals with all information you feel is important and has not been asked for already.

__Test Format:__

The test format should be reviewed according to the physical make-up of the test.

A. __Illustrations:__ This column deals with the illustrations contained within the test. Are the illustrations appropriate, clear, and do they add to the student's understanding or do they distract the student?

B. __Print Size:__ This column deals with whether the print size for items and/or directions is appropriate for the grade(s) tested.

C. __Items/Page:__ This column deals with whether the number of items per page is appropriate for the grade(s) the test level covers. Too many items per page increases difficulty unnecessarily.

D. __Item Arrangement:__ This column deals with the physical placement of the items on a page. Are the items and item directions clearly associated? Are graphs or charts necessary to answer the item(s) on the same page as the item(s)?

E. __Color:__ This column deals with whether the print color is acceptable or distracting to the student. Red, for instance, is a difficult hue for students to read.

F. __Clarity of Test Directions Within the Test:__ This column deals primarily with student test directions. Are the directions provided so that the student knows what to do for each new set of items?

G. __Sample Items:__ This column deals with the appropriateness and clarity of the examples(s) provided to students prior to each section of the test.

Test _____

Level _____

Rater _____

Key

0 - Poor/nonexistent/inappropriate

1 - Weak/somewhat inappropriate

2 - Adequate/appropriate

3 - Good/very appropriate

Test Format: Subtests	(A) Illus- trations	(B) Print Size	(C) Items/Page	(D) Item Arrangement	(E) Color	(F) Clarity of Test Dir. W/I Test	(G) Sample Items	Comments
Word Knowledge	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Reading Comp.	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
(1) Total Reading	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Spelling	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Language	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
(2) Total Language	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Computation	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Concepts	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Problem Solving	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
(3) Total Mathematics	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
Others ( ) _____	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
( ) _____	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	
( ) _____	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	

TEST REVIEW FORM (Pg. 4)

General Directions:

1. All information requested at the top of each form should be completed before starting the review process.

2. Subtests: This column contains the names of subtests normally found in standardized achievement tests. The subtests are grouped by major areas: Total Reading (1), Total Language, (2), and Total Mathematics (3). If other or additional subtests exist, please list each at the bottom in the area labeled "Other". Assign the area 1, 2, or 3 and write the title of the subtest.

3. Comments: This column deals with all information you feel is important and has not been asked for already.

Test Administration:

Test Administration should be reviewed on the basis of how it handles all areas necessary to administering the test to the students.

A. Time Allotment: This column deals with whether the amount of time given to complete a subtest is reasonable for the majority of the students.

B. Administration Directions: This column deals with whether the directions to the teacher concerning test administration are clear.

C. Information Content of Manual: This column deals with whether the administration manual is comprehensive. Does the manual adequately and usefully explain each subtest, score interpretation, norms, etc.?

Test _____

Level _____

Rater _____

Key
0 - Poor/nonexistent/inappropriate
1 - Weak/somewhat inappropriate
2 - Adequate/appropriate
3 - Good/very appropriate

Test Administration: Subtests	(A) Time Allotment	(B) Administration Directions	(C) Information Content of Manual	Comments
Word Knowledge	0 1 2 3	0 1 2 3	0 1 2 3	
Reading Comp.	0 1 2 3	0 1 2 3	0 1 2 3	
(1) Total Reading	0 1 2 3	0 1 2 3	0 1 2 3	
Spelling	0 1 2 3	0 1 2 3	0 1 2 3	
Language Arts	0 1 2 3	0 1 2 3	0 1 2 3	
(2) Total Language	0 1 2 3	0 1 2 3	0 1 2 3	
Computation	0 1 2 3	0 1 2 3	0 1 2 3	
Concepts	0 1 2 3	0 1 2 3	0 1 2 3	
Problem Solving	0 1 2 3	0 1 2 3	0 1 2 3	
(3) Total Mathematics	0 1 2 3	0 1 2 3	0 1 2 3	
Others				
_____	0 1 2 3	0 1 2 3	0 1 2 3	
_____	0 1 2 3	0 1 2 3	0 1 2 3	
_____	0 1 2 3	0 1 2 3	0 1 2 3	

TEST REVIEW FORM (Pg. 5)

General Directions:

1. All information requested at the top of each form should be completed before starting the review process.

2. Comments: This column deals with all information you feel is important and has not been asked for already.

Reporting Form:

Reports should be reviewed on the basis of the type of information that can be provided by the test to the various audiences (i.e., teachers, students, parents, administrators, etc.).

A. Yes No: This column deals with whether information exists by student, class, grade/school, other.

B. Completeness of Information: This column examines whether the information provided adequately covers all facets of the test.

C. Curriculum Decisions: This column deals with whether the information provided can be used to make curriculum decisions. Can student/class/grade strengths and weaknesses be identified?

D. Subtest Analysis: This column deals with whether the report form provides information by subtest.

E. Report Forms: This column deals with the clarity of the reporting forms. Are the forms easy to read and can the information they contain be readily located?

F. Facilitates Interpretation: This column deals with whether the report forms inhibit or enhance interpretation of test results. This column can be viewed as a cumulation of Columns A through E.

Test _____

Level _____

Rater _____

Key
0 - Poor/nonexistent/inappropriate
1 - Weak/somewhat inappropriate
2 - Adequate/appropriate
3 - Good/very appropriate

Reports:

Information Available by	(A) Yes No	(B) Completeness of Information	(C) Curriculum Decisions	(D) Subtest Analysis	(E) Report Forms	(F) Facilitates Interpretation
Student	___ ___	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Class	___ ___	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Grade/School	___ ___	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
Other Reports	___ ___	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
( ) _____	___ ___	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
( ) _____	___ ___	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3
( ) _____	___ ___	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3	0 1 2 3

COMMENTS: