DOCUMENT RESUME

ED 236 199 TM 830 708

AUTHOR Tindal, Gerald; And Others
TITLE The Technical Adequacy of a Basal Series Mastery Test: The Scott-Foresman Reading Program.
INSTITUTION Minnesota Univ., Minneapolis. Inst. for Research on Learning Disabilities.
SPONS AGENCY Office of Special Education and Rehabilitative Services (ED), Washington, DC.
REPORT NO IRLD-RR-128
PUB DATE Jul 83
CONTRACT 300-80-0622
NOTE 37p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Basal Reading; Criterion Referenced Tests; Grade 4; Intermediate Grades; *Mastery Tests; Measurement Techniques; Reading Research; *Reading Tests; *Test Reliability; *Test Validity
IDENTIFIERS *Scott Foresman Company; *SRA Basic Reading Series; Word Reading Test

ABSTRACT

The purpose of this study was to examine the reliability and validity of a basal reading series mastery test. Subjects were 25 fourth graders, who were tested once on the SRA Reading Achievement Test, twice on the Scott-Foresman End-of-Book 9 Criterion-referenced Test (CRT), and once on the Word Reading Test. Traditional psychometric correlational analyses as well as strategies specifically designed for examining the adequacy of criterion-referenced tests were applied to the data to investigate the following dimensions of the technical adequacy of the CRT: consistency of student performance across two administrations of the CRT, criterion-related validity of the CRT scores with respect to two other measures of reading proficiency, and criterion-related validity of the CRT mastery/nonmastery decisions with respect to pre/post instructional status. Results indicated that the reliability and validity was acceptable for the total test and the scale scores, with the exception of the Literacy Understanding/Appreciation scale and, in some cases, the Word Identification scale. Implications for the development and use of criterion-referenced tests are discussed. (Author)

University of Minnesota

Research Report No. 128

THE TECHNICAL ADEQUACY OF A BASAL SERIES MASTERY TEST: THE SCOTT-FORESMAN READING PROGRAM

Gerald Tindal, Lynn Fuchs, Douglas Fuchs, Mark Shinn,
Stanley Deno, and Gary Germann

iRLD

Institute for Research on Learning Disabilities

Director: James E. Ysseldyke

The Institute for Research on Learning Disabilities is supported by a contract (300-80-0622) with the Office of Special Education, Department of Education, through Title VI-G of Public Law 91-230. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled students.

During 1980-1983, Institute research focuses on four major areas:

- Referral
- Identification/Classification
- Intervention Planning and Progress Evaluation
- Outcome Evaluation

Additional information on the Institute's research objectives and activities may be obtained by writing to the Editor at the Institute (see Publications list for address).

The research reported herein was conducted under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent the official position of the Office of Special Education.

Research Report No. 128

THE TECHNICAL ADEQUACY OF A BASAL SERIES MASTERY TEST: THE SCOTT-FORESMAN READING PROGRAM

Gerald Tindal, Lynn Fuchs, Douglas Fuchs, Mark Shinn,
and Stanley L. Deno
Institute for Research on Learning Disabilities
University of Minnesota

Gary Germann
Pine County Educational Cooperative

July, 1983

Abstract

The purpose of this study was to examine the reliability and validity of a basal reading series mastery test. Subjects were 25 fourth graders, who were tested once on the SRA Reading Achievement Test, twice on the Scott-Foresman End-of-Book 9 Criterion-referenced Test (CRT), and once on the Word Reading Test. Traditional psychometric correlational analyses as well as strategies specifically designed for examining the adequacy of criterion-referenced tests were applied to the data to investigate the following dimensions of the technical adequacy of the CRT: (a) consistency of student performance across two administrations of the CRT, and (b) criterion-related validity of the CRT scores with respect to two other measures of reading proficiency and criterion-related validity of the CRT mastery/nonmastery decisions with respect to pre/post instructional status. Results indicated that the reliability and validity was acceptable for the total test and the scale scores, with the exception of the Literary Understanding/ Appreciation scale and, in some cases, the Word Identification scale. Implications for the development and use of criterion-referenced tests are discussed.

The Technical Adequacy of a Basal Series Mastery Test:

The Scott-Foresman Reading Program

Within the past decade, interest in and the use of criterion-referenced (CR) testing as a tool for evaluating the effects of instructional programs has expanded. In contrast to traditional global norm-referenced achievement tests, which typically have poor content validity with respect to classroom curricula, CR instruments are isomorphic with instructional programs and potentially useful for measuring the extent to which individuals or groups have mastered specific educational objectives.

Despite the increased focus on CR tests in the schools, there has been a lack of concomitant investigation of the reliability and validity of such instruments. Teachers who create their own CR tests rarely possess resources to conduct expensive, time-consuming reliability and validity studies. Additionally, Tindal, Shinn, Fuchs, Fuchs, Deno, and Germann (1983) documented that publishers of commercial CR instruments typically fail to address technical adequacy at all; and, when adequacy is examined, developers rely predominantly on traditional psychometric correlational analyses, which have been criticized for use with CR tests (Popham & Husek, 1969). Therefore, although CR tests may appear to be useful for evaluating the effects of instructional programs because of strong content validity, it remains unclear whether they are accurate (reliable) or meaningful (valid) for their intended purposes.

In response to this problem, researchers recently have begun the process of investigating the psychometric characteristics of commercially available basal reading series CR tests. Fuchs, Tindal,

Shinn, Fuchs, Deno, and Germann (1983) examined the test-retest reliability and criterion validity of a Ginn 720 basal mastery test and found its quality variable. In most analyses, the study skills subtests appeared inadequate, the quality of the comprehension subtests varied, and the decoding and vocabulary subtests and the total score were acceptable. Tindal et al. (1983) determined that a mastery test from the Houghton-Mifflin reading series was less than adequate; the decoding and comprehension test scales were both unreliable and invalid. These findings suggest that content and face validity are necessary but insufficient dimensions of CR test adequacy, and that test consumers must seek empirical validation of each CR test before relying on such test data for making instructional decisions.

The purpose of the current study was to extend the work of Tindal et al. (1983) and Fuchs et al. (1983) by examining the reliability and validity of another basal series mastery test, that of Scott-Foresman (1981). In doing so, the present study sought to increase the data base concerning the adequacy of CR tests in order to provide relevant information not only to consumers of this specific measure but also to users of other CR instruments for which technical data are still unavailable.

Method

Subjects

Subjects were 25 students (13 M, 12 F) from one fourth grade class, located in a school district of a rural midwestern cooperative. The students' mean reading percentile rank was 65.0 (SD 31.93) as

measured on the Science Research Associates (SRA) Reading Achievement Test. Only those students for whom there were no missing data were included in any given analysis.

Measures

Three measures of reading performance were used in the study: a basal series criterion-referenced test, a global norm-referenced test, and a curriculum-based word reading test.

Criterion-referenced test. Four scales of the End-of-Book 9 Criterion-referenced Test (CRT; Johns, 1981) of the Scott-Foresman reading series were employed as measures. Each of the four scales, Word Identification, Comprehension, Study and Research, and Literary Understanding and Appreciation, is comprised of subtests. Table 1 lists the subtests constituting each scale and provides brief descriptions of tasks the examinee is required to do within each subtest. This CRT includes between 12 and 43 items per scale and cutoff scores are established at 79% and 83% correct responses. For the purpose of this study, two subtests, Fiction and Nonfiction from the Comprehension scale and Summarizing from the Study and Research scale were omitted; they are not described in Table 1. With these omissions, items per scale ranged from 12 to 41 and the mastery cutoff scores fell between 76% and 83% correct responses.

Insert Table 1 about here

Norm-referenced test. The Science Research Associates (SRA) Reading Achievement Test (Naslund, Thorpe, & Lefever, 1978) is

comprised of two subtests: vocabulary and comprehension. In the vocabulary section, examinees are required to select, from four alternatives, a synonym for an underlined word in a sentence. In the comprehension section, examinees read 200-300 word passages and answer questions in a multiple choice format. Total test score is based on a linear combination of the two subtests. Internal consistency reliability was reported at .88 (Salvia & Ysseldyke, 1981).

Curriculum-based word reading test. The Word Reading Test (Deno, Mirkin, & Chiang, 1982) requires children to read aloud passages and isolated word lists and is scored in terms of average numbers of words correct and incorrect over two alternate forms of the Isolated Word Reading and Passage Reading scales. The 200-word passages are drawn randomly from a student's grade appropriate basal reading book; the 150-word lists sample words randomly from the basals, with 60% of words drawn from the student's grade appropriate level and 40% sampled equally from all previous levels. For the passage and isolated Word Reading Test, test-retest and alternate form reliabilities were at least .90 (Fuchs, Deno, & Marston, in press; Fuchs, Wesson, Tindal, Mirkin, & Deno, 1981).

Procedure

All students were tested in groups by a school psychologist on the SRA Reading Achievement Test, and by their classroom teachers on the CRT. The Word Reading Test was administered individually by trained aides. Standardized administration procedures were adhered to on all tests. Testing time ranged from 60 to 90 minutes for the SRA Test, 60 to 90 minutes for the CRT, and five to six minutes for the

Word Reading Test. Students were tested on the following measures in the following order within a 2-week time period: the CRT, the SRA Reading Achievement Test, the Word Reading Test, and the CRT again.

Data Analysis

Consistency of performance on two administrations of the same test. Consistency of students' performance on the CRT was assessed in three ways. First, traditional test-retest reliability was determined by correlating scores from the two administrations of the CRT. The other two analysis strategies were designed specifically for criterion-referenced measures (see Millman, 1974). In the first of these, consistency of students' scale scores was determined by (a) computing individuals' percentage correct scores on each scale for each administration of the CRT, (b) calculating for each individual his/her difference score across the two administrations of each scale, and (c) determining the percentages of examinees having each possible difference score on each scale. In the second strategy, consistency of mastery-nonmastery decisions on scales was determined by dividing the difference between observed and chance proportions of agreements in decisions by the maximum value that difference could assume. (The chance proportion of agreements was computed by multiplying and then summing the marginal proportions of the same decision categories for the two administrations, as done in a chi-square test of association.)

Criterion-related validity. The criterion-related validity of the CRT was determined in two ways. The traditional psychometric strategy of correlating scores on the measure of interest (CRT) with criterion measures was used. The SRA Reading Achievement Test and the

Word Reading Test were employed as the criterion measures. Additionally, chi-square statistical tests were applied to contingency tables wherein test mastery-nonmastery represented one dimension of each table and pre-post instructional status represented the other dimension. Percentages of misclassifications and ϕ coefficients supplemented the chi-square tests.

Results

Table 2 is a display of students' mean scores and standard deviations on each scale and for the total score of the CRT, on the subtest and total scores of the SRA Reading Achievement Test, and on the isolated word reading and passage reading scales of the Word Reading Test.

Insert Table 2 about here

Consistency of Performance on Administrations of the Same Test

Test-retest reliability correlations on scales of the CRT are displayed in Table 3. All coefficients for the test scales and total test were at least .90, with the exception of the Literary Understanding/Appreciation scale, which had a coefficient of .68.

Insert Table 3 about here

The second analysis of the consistency of performance involved calculating the percentages of examinees who had different percentage correct scores across the two administrations of the CRT. Figures 1

and 2 are graphic displays of the percentages of examinees displaying various difference scores on each scale of the CRT; Table 4 summarizes the information illustrated on the graphs. The range of difference scores on the scales fell between 0 and 42%. The percentage of examinees with 0% difference scores on two administrations ranged from 13 on the Comprehension scale to 25 on the Literary Understanding/Appreciation scale.

Insert Figures 1-2 and Table 4 about here

The third analysis of the consistency of performance addressed consistency of mastery-nonmastery decisions across the two administrations of the CRT. Table 5 is a display of the uncorrected and corrected proportions of examinees placed into the same decision category on the two administrations. On the Word Identification, Comprehension, and Study and Research scales, the corrected proportions were high, ranging from a proportion of agreement on the Study and Research scale of 84% higher than chance to a proportion of agreement on the Word Identification scale of 96% greater than chance. On the Literary Understanding/Appreciation scale, the proportion of agreement was lower, with a proportion of agreement of 57% better than chance.

Insert Table 5 about here

Criterion-related Validity

Correlational analyses were conducted between the CRT scales and two criterion measures, the SRA Reading Achievement Test and the Word Reading Test. Correlations between the CRT scales and the SRA subtest and total test scores are displayed in Table 6. They ranged from .48 to .90 when CRT scale and SRA vocabulary subtest scores were involved; from .59 to .87 when CRT scale and SRA comprehension subtest scores were employed; and from .55 to .89 when CRT scale and SRA total scores were used. For the CRT total score, correlations ranged between .92 and .95. The median correlation for the CRT Word Identification scale was .62; for the CRT Comprehension scale, .86; for the CRT Study and Research scale, .89; and for the Literary Understanding/Appreciation scale, .55.

Insert Table 6 about here

Correlations between the CRT scales and the Word Reading Test scale scores are displayed in Table 7. They ranged from .42 to .73 when isolated word reading scores and CRT scale scores were involved, and from .55 to .76 when passage reading scores and CRT scale scores were employed. For the CRT total score, correlations were .77 and .84.

Insert Table 7 about here

Criterion validity also was examined by inspecting the relation between mastery-nonmastery decisions on the CRT and actual pre-post

instructional status. Relevant chi-square values, phi coefficients, and percentages of misclassified students are displayed in Table 8. The highest percentages of misclassifications occurred with the Word Identification (48%) and Literary Understanding/Appreciation (35%) scales; lower percentages were found for the Study and Research scale and the total score (22%). The percentage of misclassified students on the Comprehension scale was 26.

Insert Table 8 about here

Discussion

The purpose of the current study was to describe the reliability and validity of a basal reading series criterion-referenced mastery test. The study examined two aspects of the technical adequacy of the Scott-Foresman End-of-Book 9 Criterion-referenced test: (a) the consistency of students' performance on two administrations of the test, and (b) the criterion validity of the test with respect to two other measures of reading proficiency that have demonstrated psychometric strength. On these indices, the total score and all scale scores, with the exception of the Literary Understanding scale and, in some cases, the Word Identification scale, seemed adequate.

Test-retest reliability coefficients indicated that, when the CRT was administered twice within a short time interval, students' performance was somewhat inconsistent on the Literary Understanding and Appreciation scale, with the correlation falling below the acceptable range for making even group decisions (Salvia & Ysseldyke,

1981). Nevertheless, for the remaining scales and the total score, correlations were high and fell into the acceptable range for individual decision making.

The pattern of results of this traditional correlational analysis of consistency of student performance across testings was corroborated by the criterion-referenced strategy of examining the proportions of examinees consistently classified into the same decision category. As with the correlational analyses, statistics were generally high. All corrected proportions fell above 84% better than chance agreement except the Literary Understanding/Appreciation scale score, for which the corrected proportion fell below 60% better than chance.

Inspection of the consistency of test scores displayed in Figure 1 and 2 and in Table 4 reveals that results of the second criterion-referenced strategy for examining test consistency also corroborated the correlational results. The average percentage of subjects obtaining the same score across all scales was 18.0. Interestingly, the percentage for the Literary Understanding/Appreciation scale was higher than the percentages for the remaining scales. Nevertheless, for this Literary Appreciation/Understanding scale, greater percentages of subjects also scored with relatively great discrepancy, and the 37% of subjects whose differences scores were between 15 and 42% across the two testings achieved scores sufficiently discrepant to result in numerous inconsistent mastery decisions.

The criterion validity of the CRT also was examined in this study. The traditional correlational analyses indicated that the criterion validity of the CRT scale scores with respect to the SRA

Reading Achievement Test was good, with 50% of correlations between the CRT scales and the SRA subtests falling above .80. With the Word Reading Test, correlations between the CRT scales and the Word Reading Test scales were generally lower, with 50% falling above .70 and no correlations above .80.

With respect to both the SRA and the Word Reading Tests, correlations based on the Word Identification and the Literary Understanding/Appreciation scales typically were relatively low. The finding that criterion validity with the SRA Test was greater than with the Word Reading Test is contrary to previous findings (Fuchs et al., 1983). Given that both the Word Reading Test and the CRT are curriculum-based, one might expect a stronger relation between these two measures, and current findings are surprising. Nevertheless, correlations among curriculum-based measures and more global indices, such as the SRA test, have been reported frequently at high levels (Fuchs et al., in press; Fuchs, Fuchs, & Deno, 1982), and correlations between the CRT and the SRA tests are comparable to those reported earlier. Further, performance on the CRT predicts concurrent performance on more global measures of reading proficiency better than other basal mastery tests that have been examined (Fuchs et al., 1983; Tindal et al., 1983).

The criterion validity of the CRT also was investigated with the criterion-referenced strategy of examining the relation between the mastery-nonmastery classification on the CRT and actual pre-post instructional status. Percentages of misclassifications ranged from 22% to 48% on the CRT scales, with 22% of students misclassified on

the total test. These figures suggest that, for classifying students into groups for instruction within the basal reader for which the CRT was designed, the Word Identification and Literary Understanding/ Appreciation scales have limited utility whereas the Comprehension and Study and Research scales as well as total test score are more valid.

Consequently, the current study suggests that the Scott-Foresman End-of-Book 9 CRT is generally of good quality. Much of the test seemed useful both for predicting global reading proficiency and for making decisions about student placement and progress within the curriculum. Specifically, in the test-retest consistency analyses the total score and the scale scores, with the exception of the Literary Understanding/Appreciation scale, appeared adequate. In the criterion validity analyses, the total and the scale scores, with the exception of the Word Identification and the Literary Understanding/Appreciation scales, demonstrated technical strength. This indicates that (a) educators should use this CRT judiciously, relying primarily on the Comprehension, Study and Research, and total scores for making decisions about student performance and mastery in the curriculum, and (b) test developers at Scott-Foresman might consider reexamining the Word Identification and Literary Understanding/Appreciation scales.

In any case, the technical adequacy of the Scott-Foresman End-of-Book 9 CRT was superior to previously examined basal mastery tests (see Fuchs et al., 1983; Tindal et al., 1983). A probable explanation for this superiority is as follows: Whereas other CR tests that have been examined have teachers score and compute mastery scores on subtests as well as scale scores, the Scott-Foresman CRT

limits computation of mastery scores to the test scale and total scores. By doing so, the Scott-Foresman CRT requires educators to rely on information summarized across a relatively large sampling of student behavior. As the Spearman-Brown formula indicates, when the number of items in a test increases, the reliability and validity of test scores improve correspondingly. It appears that the author of this Scott-Foresman CRT has capitalized on this measurement phenomenon by eliminating the calculation and consideration of mastery scores based on subtests, which incorporate relatively few items.

Finally, in this study, results based on traditional and on criterion-referenced strategies for examining test adequacy were analogous. Wherever traditional correlational statistics suggested relative strengths or weaknesses in scales, the results of the alternative strategies paralleled findings. Consequently, results of the present study echo previous research (Tindal et al., 1983), which suggests that the two types of analyses corroborate, complement, and enhance each other. It appears that both strategies may be appropriate and necessary for investigating and describing the reliability and validity of criterion-referenced tests.

References

Deno, S. L., Mirkin, P. K., & Chiang, B. Identifying valid measures of reading. Exceptional Children, 1982, 49(1), 36-45.

Fuchs, L. S., Deno, S. L., & Marston, D. Improving the reliability of curriculum-based measures of academic skills for psychoeducational decision making. Diagnostique, in press.

Fuchs, L. S., Fuchs, D., & Deno, S. L. Reliability and validity of curriculum-based informal reading inventions. Reading Research Quarterly, 1982, 18(1), 6-26.

Fuchs, L. S., Tindal, G., Shinn, M., Fuchs, D., Deno, S. L., & Germann, G. The technical adequacy of a basal series mastery test: The Ginn 720 Series (Research Report No. 122). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1983.

Fuchs, L. S., Wesson, C., Tindal, G., Mirkin, P. K., & Deno, S. L. Teacher efficiency in continuous evaluation of IEP goals (Research Report No. 53). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1981.

Johns, J. Scott, Foresman Reading: Sea Treasures End-of-Book Test. Glenview, Illinois: Scott, Foresman, 1981.

Millman, J. Criterion-referenced measurement. In W. J. Popham (Ed.), Evaluation in education: Current applications. Berkeley: McCutchan, 1974.

Naslund, R. A., Thorpe, L. P., & Lefever, D. W. SRA achievement series: Reading, mathematics, and language arts. Chicago: Science Research Associates, 1978.

Popham, W. J., & Húsek, T. R. Implications of criterion-referenced measurement. Journal of Educational Measurement, 1969, 6, 1-9.

Salvia, J., & Ysseldyke, J. E. Assessment in special and remedial education (2nd ed.). Boston: Houghton-Mifflin, 1981.

Scott-Foresman reading. Glenview, IL: Scott, Foresman, 1981.

Tindal, G., Shinn, M., Fuchs, L., Fuchs, D., Deno, S., & Germann, G. The technical adequacy of a basal series reading mastery test (Research Report No. 113). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1983.

Table 1

Examinees' Tasks on the Scott-Foresman End-of-Book 9 Test

Scale/Subtest	Examinees' Tasks
Word Identification	
Context and Consonants	Read a sentence containing a word from which all vowels and vowel teams have been deleted. From an array of three choices, select a complete word to substitute for the incomplete one.
Syllables, Accent	Read questions concerning the syllable division and accented syllable of a word. From two choices, select the answer to each question.
Compounds, Contractions	From an array of three choices, select either a compound word or a contraction, as directed.
Root words, Ending, Suffixes	1. Given a word and three possible roots, select the correct one. 2. Given a choice of five descriptions concerning what might happen to a root word before an ending is added and given a word with an ending, select the correct description.
Comprehension	
Unfamiliar Words	Read a paragraph containing an underlined word, and select from among three choices a synonomous word or phrase.
Idioms	Read a paragraph containing an underlined idiom. From an array of three choices, select a phrase that defines the idiom.
Analogous Relationships	Read a story and (a) select, from an array of three choices, a description of how two objects described in the paragraph are alike, and (b) select, from an array of three choices, a word or noun phrase that completes a sentence describing an analogy.

Table 1 (continued)

Scale/Subtest	Examinees' Tasks
Story Problem/Solution	Read a story and answer multiple choice questions concerning the story's content.
Main Idea/Supporting Details	Read a short article. Determine whether sentences taken from that article are main ideas, supporting details, or neither.
Study and Research	
Table of Contents	Given tables of content from two books, answer multiple choice questions concerning (a) the content of the books and (b) how to access information from them.
Index	Given a partial index from a book, use it to answer multiple choice questions concerning how to access information from the book.
Encyclopedia	Given an illustration of a 21-volume encyclopedia, answer multiple choice questions concerning how to access information from the encyclopedia.
Footnotes	Read a segment from a factual article that contains footnotes, answer multiple choice questions concerning the content and use of the footnotes.
Headnotes	Read a headnote and then answer multiple choice questions concerning the content of the headnote.
Classifying	From an array of four words, select the one that does not belong with the others.
Diagrams	Given a diagram of a ship, answer multiple choice questions concerning the layout and contents of the ship.
Literary Understanding/Appreciation	
Story Elements	Read a story and answer multiple choice questions concerning the story's content.
Elements of Style	Read a paragraph and answer yes-no questions concerning whether certain sentences in the paragraph are exaggerations.

Table 1 (continued)

Scale/Subtest	Examinees' Task
Elements of Style	Read a paragraph for which each sentence is identified with a number. Then, answer multiple choice questions concerning (a) which sentence is a flashback, (b) where personification is contained in the paragraph, and (c) the point of view with which the paragraph is written.
Types of Literature	Read a story and select from an array of three choices, the literary form of the story.

Table 2

Student Performance on Measures of Reading Achievement

Test	Number of Items	Mean	SD
End-of-Book 9 Test (N=25)			
Word Identification	36	29.8	5.6
Comprehension	25	18.3	5.3
Study and Research	41	30.8	8.5
Literary Understanding/ Appreciation	12	8.2	2.2
Total	114	86.9	19.3
SRA Reading Achievement Test (N=22)			
Vocabulary		28.5	9.1
Comprehension		34.5	11.2
Total		63.5	20.1
Word Reading Test			
Isolated Word Reading (N=22)		38.2	19.8
Passage Reading (N=26)		79.2	47.1

Table 3

Test-retest Reliabilities for Scott-Foresman End-of-Book 9 Test (N=25)

Scale	Reliability
Word Identification	.93
Comprehension	.92
Study and Research	.93
Literary Understanding/Appreciation	.68
Total	.98

Table 4

Proportion of Subjects with Varying Percentages of Difference Scores Across Two Administrations of the End-of-Book 9 Test (N=24)

Scale	N[a]	Percentage of Difference Score					
		0.0	1.0 to 7.0	8.0 to 14.0	15.0 to 24.0	25.0 to 34.0	35.0 to 44.0
Word Identification	36	17	55	28	0	0	0
Comprehension	25	13	50	29	8	0	0
Study and Research	41	17	67	4	12	0	0
Literary Understanding/ Appreciation	12	25	0	38	8	25	4

[a]Number of items on the test.

Table 5

Uncorrected and Corrected Proportions of Examinees (N=24) Placed Into the Same Decision Categories on Two Administrations of the End-of-Book 9 Test

Scale	Proportion of Examinees	
	Uncorrected	Corrected for Chance Agreement[a]
Word Identification	.96	.96
Comprehension	.92	.9
Study and Research	.83	
Literary Understanding/Appreciation	.63	.57

[a]Observed--Chance Proportions/Maxiumum Value that (Observed-Chance Proportions) Can Assume.

Table 6

Correlations Between End-of-Book 9 Test and SRA Test Scores (N=21)

Scott-Foresman Scale	SRA Vocabulary	SRA Comprehension	SRA Total
Word Identification	.57	.62	.62
Comprehension	.80	.86	.86
Study and Research	.90	.87	.89
Literary Understanding/ Appreciation	.48	.59	.55
Total Test	.92	.94	.95

Table 7

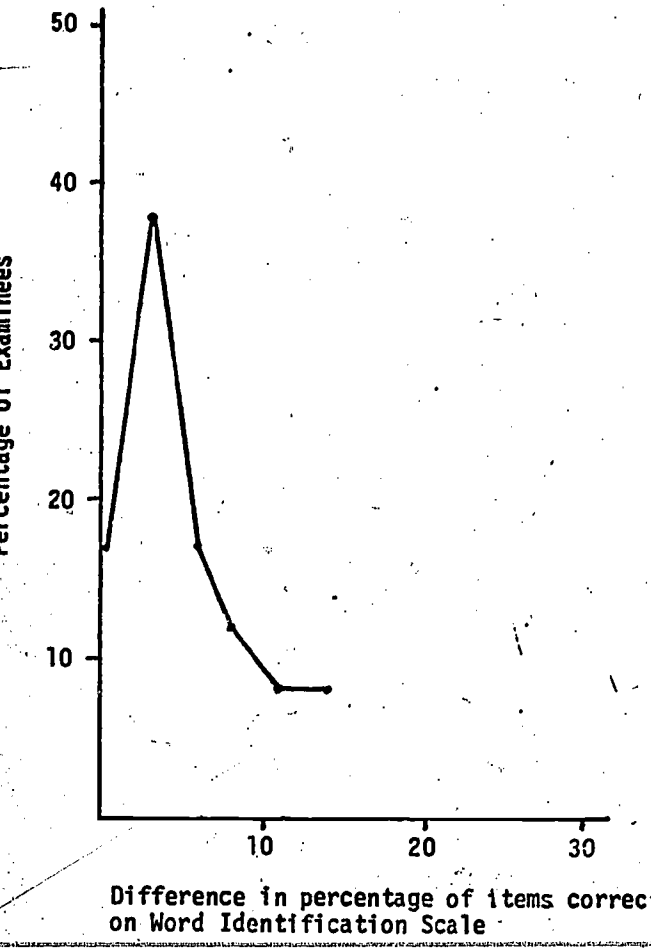
Correlations Between End-of-Book 9 Test and Word Reading Test Scores (N=21)

Scott-Foresman Scale	Word Reading Test	
	Isolated Words	Passages
Word Identification	.42	.70
Comprehension	.52	.70
Study and Research	.73	.76
Literary Understanding/Appreciation	.58	.55
Total	.77	.84

Table 8

Relation Between End-of-Book 9 Test and Criterion Classification (N=23)

Scale	χ^2	Ø	Percentage Misclassified
Word Identification	.68	.17	48
Comprehension	6.88	.55	26
Study and Research	9.79	.65	22
Literary Understanding/Appreciation	.52	.15	35
Total	9.79	.65	22



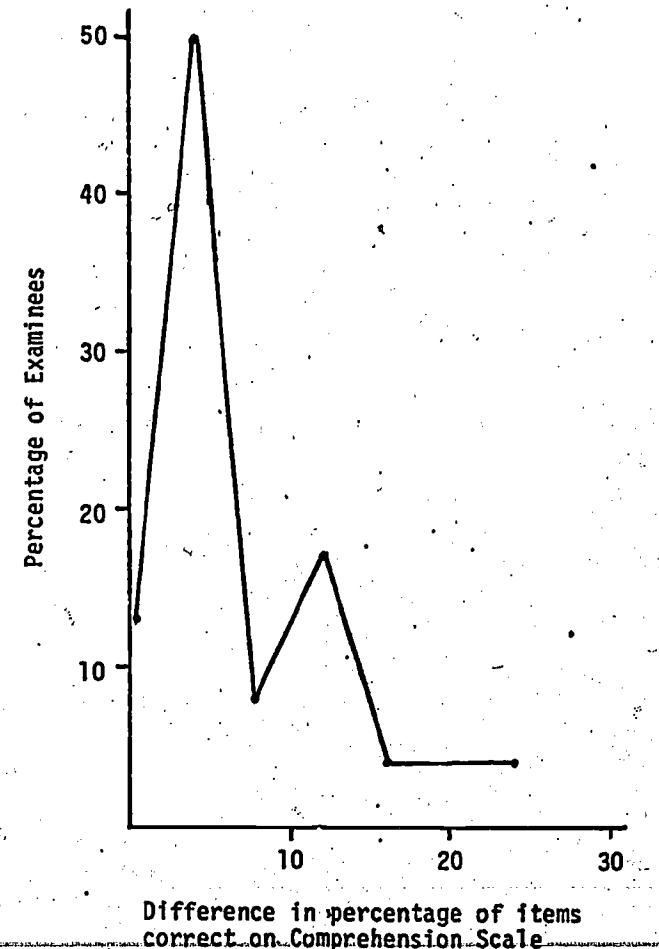
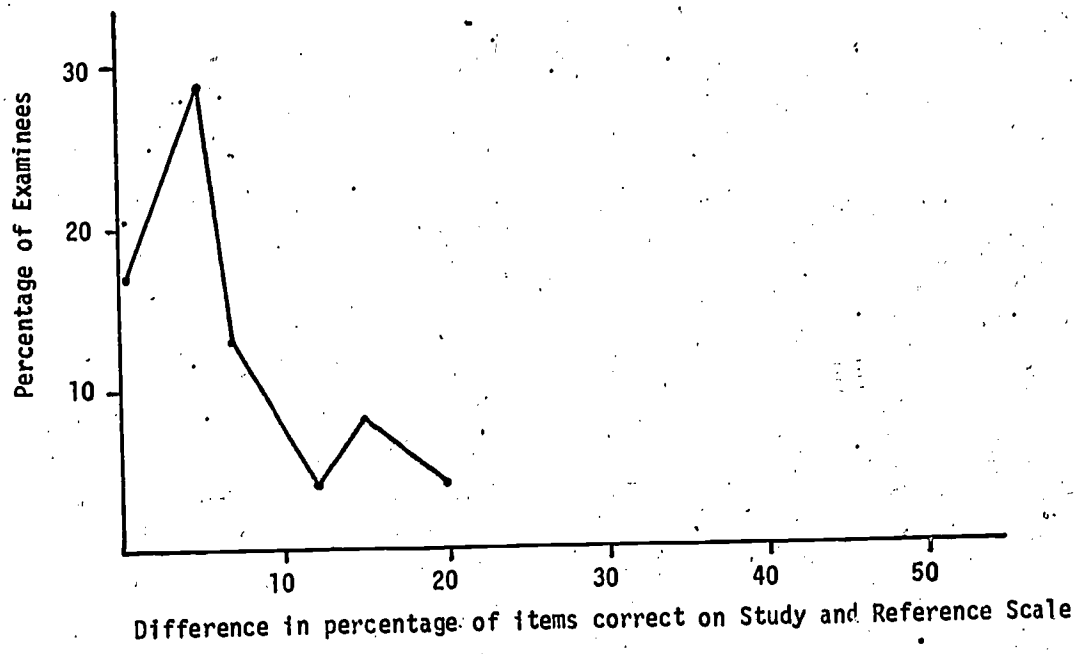


Figure 1. Displays of consistency of test scores on Word Identification and Comprehension Scales of the End-of-Book 9 CRT.



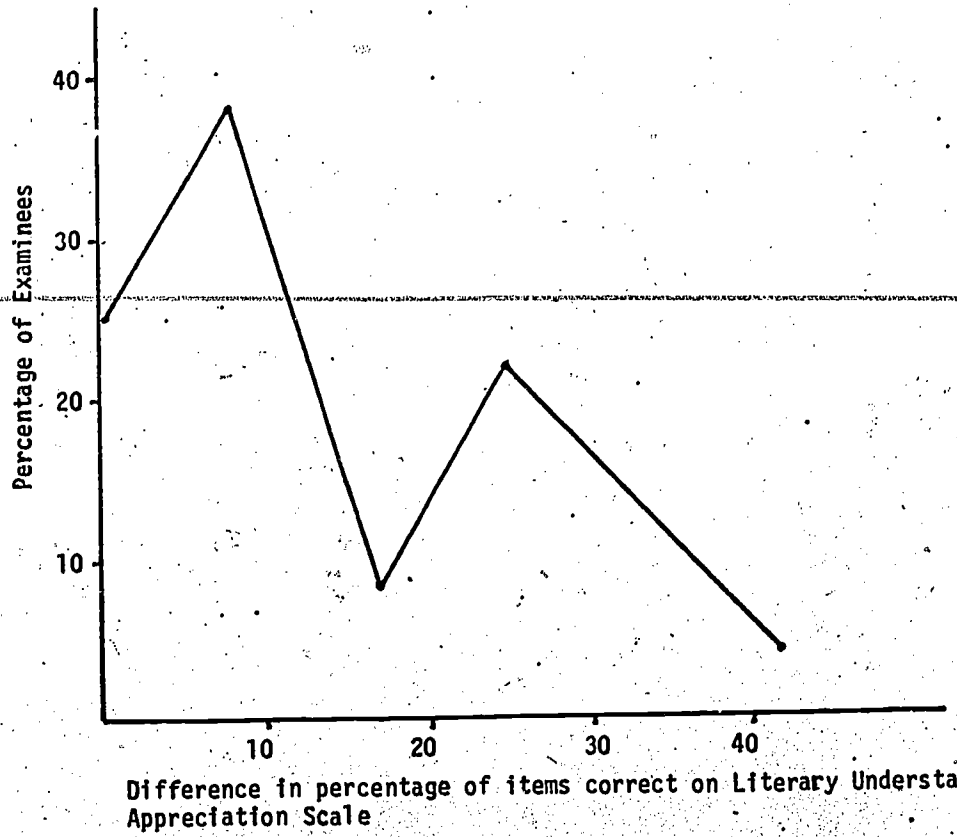


Figure 2. Displays of consistency of test scores on Study and Reference and Literary Understanding/Appreciation Scales on the End-of-Book 9 CRT.

PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for $4.00 each, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid. Requests should be directed to: Editor, IRLD, 350 Elliott Hall; 75 East River Road, University of Minnesota, Minneapolis, MN 55455.

The publications listed here are only those that have been prepared since 1982. For a complete, annotated list of all IRLD publications, write to the Editor.

Wesson, C., Mirkin, P., & Deno, S. Teachers' use of self instructional materials for learning procedures for developing and monitoring progress on IEP goals (Research Report No. 63). January, 1982.

Fuchs, L., Wesson, C., Tindal, G., Mirkin, P., & Deno, S. Instructional changes, student performance, and teacher preferences: The effects of specific measurement and evaluation procedures (Research Report No. 64). January, 1982.

Potter, M., & Mirkin, P. Instructional planning and implementation practices of elementary and secondary resource room teachers: Is there a difference? (Research Report No. 65). January, 1982.

Thurlow, M. L., & Ysseldyke, J. E. Teachers' beliefs about LD students (Research Report No. 66). January, 1982.

Graden, J., Thurlow, M. L., & Ysseldyke, J. E. Academic engaged time and its relationship to learning: A review of the literature (Monograph No. 17). January, 1982.

King, R., Wesson, C., & Deno, S. Direct and frequent measurement of student performance: Does it take too much time? (Research Report No. 67). February, 1982.

Greener, J. W., & Thurlow, M. L. Teacher opinions about professional education training programs (Research Report No. 68). March, 1982.

Algozzine, B., & Ysseldyke, J. Learning disabilities as a subset of school failure: The oversophistication of a concept (Research Report No. 69). March, 1982.

Fuchs, D., Zern, D. S., & Fuchs, L. S. A microanalysis of participant behavior in familiar and unfamiliar test conditions (Research Report No. 70). March, 1982.

Shinn, M. R., Ysseldyke, J., Deno, S., & Tindal, G. A comparison of psychometric and functional differences between students labeled learning disabled and low achieving (Research Report No. 71). March, 1982.

Thurlow, M. L. Graden, J., Greener, J. W., & Ysseldyke, J. E. Academic responding time for LD and non-LD students (Research Report No. 72). April, 1982.

Graden, J., Thurlow, M., & Ysseldyke, J. Instructional ecology and academic responding time for students at three levels of teacher-perceived behavioral competence (Research Report No. 73). April, 1982.

Algozzine, B., Ysseldyke, J., & Christenson, S. The influence of teachers' tolerances for specific kinds of behaviors on their ratings of a third grade student (Research Report No. 74). April, 1982.

Wesson, C., Deno, S., & Mirkin, P. Research on developing and monitoring progress on IEP goals: Current findings and implications for practice (Monograph No. 18). April, 1982.

Mirkin, P., Marston, D., & Deno, S. L. Direct and repeated measurement of academic skills: An alternative to traditional screening, referral, and identification of learning disabled students (Research Report No. 75). May, 1982.

Algozzine, B., Ysseldyke, J., Christenson, S., & Thurlow, M. Teachers' intervention choices for children exhibiting different behaviors in school (Research Report No. 76). June, 1982.

Tucker, J., Stevens, L. J., & Ysseldyke, J. E. Learning disabilities: The experts speak out (Research Report No. 77). June, 1982.

Thurlow, M. L., Ysseldyke, J. E., Graden, J., Greener, J. W., & Mecklenberg, C. Academic responding time for LD students receiving different levels of special education services (Research Report No. 78). June, 1982.

Graden, J. L., Thurlow, M. L., Ysseldyke, J. E., & Algozzine, B. Instructional ecology and academic responding time for students in different reading groups (Research Report No. 79). July, 1982.

Mirkin, P. K., & Potter, M. L. A survey of program planning and implementation practices of LD teachers (Research Report No. 80). July, 1982.

Fuchs, L. S., Fuchs, D., & Warren, L. M. Special education practice in evaluating student progress toward goals (Research Report No. 81). July, 1982.

Kuehnle, K., Deno, S. L., & Mirkin, P. K. Behavioral measurement of social adjustment: What behaviors? What setting? (Research Report No. 82). July, 1982.

Fuchs, D., Dailey, Ann Madsen, & Fuchs, L. S. Examiner familiarity and the relation between qualitative and quantitative indices of expressive language (Research Report No. 83). July, 1982.

Videen, J., Deno, S., & Marston, D. Correct word sequences: A valid indicator of proficiency in written expression (Research Report No. 84). July, 1982.

Potter, M. L. Application of a decision theory model to eligibility and classification decisions in special education (Research Report No. 85). July, 1982.

Greener, J. E., Thurlow, M. L., Graden, J. L., & Ysseldyke, J. E. The educational environment and students' responding times as a function of students' teacher-perceived academic competence (Research Report No. 86). August, 1982.

Deno, S., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study (Research Report No. 87). August, 1982.

Skiba, R., Wesson, C., & Deno, S. L. The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison (Research Report No. 88). September, 1982.

Marston, D., Tindal, G., & Deno, S. L. Eligibility for learning disability services: A direct and repeated measurement approach (Research Report No. 89). September, 1982.

Thurlow, M. L., Ysseldyke, J. E., & Graden, J. L. LD students' active academic responding in regular and resource classrooms (Research Report No. 90). September, 1982.

Ysseldyke, J. E., Christenson, S., Pianta, R., Thurlow, M. L., & Algozzine, B. An analysis of current practice in referring students for psychoeducational evaluation: Implications for change (Research Report No. 91). October, 1982.

Ysseldyke, J. E., Algozzine, B., & Epps, S. A logical and empirical analysis of current practices in classifying students as handicapped (Research Report No. 92). October, 1982.

Tindal, G., Marston, D., Deno, S. L., & Germann, G. Curriculum differences in direct repeated measures of reading (Research Report No. 93). October, 1982.

Fuchs, L.S., Deno, S. L., & Marston, D. Use of aggregation to improve the reliability of simple direct measures of academic performance (Research Report No. 94). October, 1982.

Ysseldyke, J. E., Thurlow, M. L., Mecklenburg, C., & Graden, J. Observed changes in instruction and student responding as a function of referral and special education placement (Research Report No. 95). October, 1982.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. Effects of frequent curriculum-based measurement and evaluation on student achievement and knowledge of performance: An experimental study (Research Report No. 96). November, 1982.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. Direct and frequent measurement and evaluation: Effects on instruction and estimates of student progress (Research Report No. 97). November, 1982.

Tindal, G., Wesson, C., Germann, G., Deno, S. L., & Mirkin, P. K. The Pine County model for special education delivery: A data-based system (Monograph No. 19). November, 1982.

Epps, S., Ysseldyke, J. E., & Algozzine, B. An analysis of the conceptual framework underlying definitions of learning disabilities (Research Report No. 98). November, 1982.

Epps, S., Ysseldyke, J. E., & Algozzine, B. Public-policy implications of different definitions of learning disabilities (Research Report No. 99). November, 1982.

Ysseldyke, J. E., Thurlow, M. L., Graden, J. L., Wesson, C., Deno, S. L., & Algozzine, B. Generalizations from five years of research on assessment and decision making (Research Report No. 100). November, 1982.

Marston, D., & Deno, S. L. Measuring academic progress of students with learning difficulties: A comparison of the semi-logarithmic chart and equal interval graph paper (Research Report No. 101). November, 1982.

Beattie, S., Grise, P., & Algozzine, B. Effects of test modifications on minimum competency test performance of third grade learning disabled students (Research Report No. 102). December, 1982

Algozzine, B., Ysseldyke, J. E., & Christenson, S. An analysis of the incidence of special class placement: The masses are burgeoning (Research Report No. 103). December, 1982.

Marston, D., Tindal, G., & Deno, S. L. Predictive efficiency of direct, repeated measurement: An analysis of cost and accuracy in classification (Research Report No. 104). December, 1982.

Wesson, C., Deno, S., Mirkin, P., Sevcik, B., Skiba, R., King, R., Tindal, G., & Maruyama, G. Teaching structure and student achievement effects of curriculum-based measurement: A causal (structural) analysis (Research Report No. 105). December, 1982.

Mirkin, P. K., Fuchs, L. S., & Deno, S. L. (Eds.). Considerations for designing a continuous evaluation system: An integrative review (Monograph No. 20). December, 1982.

Marston, D., & Deno, S. L. Implementation of direct and repeated measurement in the school setting (Research Report No. 106). December, 1982.

Deno, S. L., King, R., Skiba, R., Sevcik, B., & Wesson, C. The structure of instruction rating scale (SIRS): Development and technical characteristics (Research Report No. 107). January, 1983.

Thurlow, M. L., Ysseldyke, J. E., & Casey, A. Criteria for identifying LD students: Definitional problems exemplified (Research Report No. 108). January, 1983.

Tindal, G., Marston, D., & Deno, S. L. The reliability of direct and repeated measurement (Research Report No. 109). February, 1983.

Fuchs, D., Fuchs, L. S., Dailey, A. M., & Power, M. H. Effects of pre-test contact with experienced and inexperienced examiners on handicapped children's performance (Research Report No. 110). February, 1983

King, R. P., Deno, S., Mirkin, P., & Wesson, C. The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison (Research Report No. 111). February, 1983.

Tindal, G., Deno, S. L., & Ysseldyke, J. E. Visual analysis of time series data: Factors of influence and level of reliability (Research Report No. 112). March, 1983.

Tindal, G, Shinn, M., Fuchs, L., Fuchs, D., Deno, S., & Germann, G. The technical adequacy of a basal reading series mastery test (Research Report No. 113). April, 1983.

Sevcik, B., Skiba, R., Tindal, G., King, R., Wesson, C., Mirkin, P., & Deno, S. Communication of IEP goals and student progress among parents, regular classroom teachers, and administrators using systematic formative evaluation (Research Report No. 114). April, 1983.

Wesson, C. Two student self-management techniques applied to data-based program modification (Research Report No. 115). April, 1983.

Wesson, C., Skiba, R., Sevcik, B., King, R., Tindal, G., Mirkin, P., & Deno, S. The impact of the structure of instruction and the use of technically adequate instructional data on reading improvement (Research Report No. 116). May, 1983.

Wesson, C. Teacher vs student selection of instructional activities (Research Report No. 117). May, 1983.

Tindal, G., & Deno, S. Factors influencing the agreement between visual and statistical analyses of time series data (Research Report No. 118). June, 1983.

Skiba, R. S. Classroom behavior management: A review of the literature (Monograph No. 21), June, 1983.

Graden, J. L., Thurlow, M. L., & Ysseldyke, J. E. When are students most academically engaged? Academic responding time in different instructional ecologies (Research Report No. 119). June, 1983.

Fuchs, L. S., Deno, S. L., & Roettger, A. The effect of alternative data-utilization rules on spelling achievement: An n of 1 study (Research Report No. 120). June, 1983.

Skiba, R., Sevcik, B., Wesson, C., King, R., & Deno, S. The non-effect of process-product variables in resource classrooms (Research Report No. 121). June, 1983.

Fuchs, L. Tindal, G., Fuchs, D., Deno, S., & Germann, G. Technical adequacy of basal readers' mastery tests: The Ginn 720 series (Research Report No. 122). June, 1983.

Tindal, G., Germann, G., Marston, D., & Deno, S. The effectiveness of special education: A direct measurement approach (Research Report No. 123). June, 1983.

Sevcik, B., Skiba, R., Tindal, G., King, R., Wesson, C., Mirkin, P., & Deno, S. Curriculum-based measurement: Effects on instruction, teacher estimates of student progress, and student knowledge of performance (Research Report No. 124). July, 1983.

Skiba, R., Marston, D., Wesson, C., Sevcik, B., & Deno, S. L. Characteristics of the time-series data collected through curriculum-based reading measurement (Research Report No. 125). July, 1983.

Ysseldyke, J., Christenson, S., Graden, J., & Hill, D. Practical implications of research on referral and opportunity to learn (Monograph No. 22). July, 1983.

Marston, D., Deno, S., & Tindal, G. A comparison of standardized achievement tests and direct measurement techniques in measuring pupil progress (Research Report No. 126). July, 1983.

Fuchs, D., Fuchs, L. S., Tindal, G., & Deno, S. L. Variability of performance: A "signature" characteristic of learning disabled children? (Research Report No. 127). July, 1983.

Tindal, G., Fuchs, L., Fuchs, D., Shinn, M., Deno, S., & Germann, G. The technical adequacy of a basal series mastery test: The Scott-Foresman reading program (Research Report No. 128). July, 1983.

Fuchs, L. S., Fuchs, D., & Deno, S. L. The nature of inaccuracy among readability formulas (Research Report No. 129). July, 1983.