DOCUMENT RESUME

ED 236 195 TM 830 704

AUTHOR Fuchs, Lynn; And Others
TITLE Technical Adequacy of Basal Readers' Mastery Tests: The Ginn 720 Series.
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-122
PUB DATE Jun 83
CONTRACT 300-80-0622
NOTE 36p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Basal Reading; Criterion Referenced Tests; Grade 5; Intermediate Grades; *Mastery Tests; Measurement Techniques; Reading Research; *Reading Tests; *Test Reliability; *Test Validity
IDENTIFIERS *Ginn Reading 720 Series; SRA Achievement 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 47 fifth graders, who were tested on the SRA Reading Achievement Test, the Ginn 720 End-of-level 11 Mastery Test (MT), and the Word Reading Test. A subgroup of 22 children was tested a second time on the MT. 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 MT: (1) consistency of student performance across two administrations of the MT, and (2) criterion-related validity of the MT scores with respect to two other measures of reading proficiency, and criterion-related validity of the MT mastery/nonmastery decisions with respect to pre/post instructional status. Results indicated that the reliability and validity was acceptable for the composite test scores, but variable for the subtests. Implications for the development and use of criterion-referenced tests are discussed. (Author)

University of Minnesota

Research Report No. 122

TECHNICAL ADEQUACY OF BASAL READERS' MASTERY TESTS: THE GINN 720 SERIES

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

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 Special Education Programs, Department of Education. 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).

Research Report No. 122

TECHNICAL ADEQUACY OF BASAL READERS' MASTERY TESTS: THE GINN 720 SERIES

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

Gary Germann
Pine County Educational Cooperative

June, 1983

Abstract

The purpose of this study was to examine the reliability and validity of a basal reading series mastery test. Subjects were 47 fifth graders, who were tested on the SRA Reading Achievement Test, the Ginn 720 End-of-level 11 Mastery Test (MT), and the Word Reading Test. A subgroup of 22 children was tested a second time on the MT. 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 MT: (a) consistency of student performance across two administrations of the MT, and (b) criterion-related validity of the MT scores with respect to two other measures of reading proficiency and criterion-related validity of the MT mastery/nonmastery decisions with respect to pre/post instructional status. Results indicated that the reliability and validity was acceptable for the composite test scores, but variable for the subtests. Implications for the development and use of criterion-referenced tests are discussed.

Technical Adequacy of Basal Readers' Mastery Tests:

The Ginn 720 Series

Two measurement formats currently are widely used in educational settings for evaluating student progress and the effects of instructional programs. The first approach is based on the long-standing tradition of administering published, norm-referenced tests on a pre and post basis. Frequently, the tests themselves demonstrate strong psychometric characteristics such as reliability, criterion-related validity, and appropriate norms. Nevertheless, this traditional assessment practice has been criticized severely for a number of reasons, including the assessment of global versus specific skills and lack of content validity. Additionally, traditional procedures are plagued with problems of gain scores, regression toward the mean, and the poor reliability of difference scores. In an attempt to ameliorate many of these problems, particularly those of content validity, the second approach, criterion-referenced (CR) instruments, has been developed to determine a student's mastery of specific curricula. Ideally, these tests measure exactly what has been taught. Despite the strong content validity of many CR tests, there is scant research addressing the reliability and criterion-related validity of these instruments. Therefore, neither available format can be used with certainty to assess a student's progress or mastery.

In response to this dilemma, researchers recently have begun the task of investigating the validity and reliability of available CR tests. Tindal, Shinn, Fuchs, Fuchs, Deno, and Germann (1983) examined a typical mastery test from the Houghton-Mifflin reading series and

found that both its test-retest reliability and its criterion-related validity were less than adequate for the decoding and comprehension test scales. This finding documents the notion that content validity is a necessary, but insufficient aspect of criterion-referenced test adequacy, and it underscores the importance of investigating the reliability and validity of each criterion-referenced test.

The purpose of the current study was to extend the work of Tindal et al. (1983) by examining the reliability and validity of another basal series mastery test, that of Ginn 720. In doing so, the present study sought to provide information of interest not only to consumers of this specific measure but also to users of other CR tests for which technical data also are still unavailable.

Method

Subjects

Subjects were 47 students (27 M, 20 F) from two fifth grade classes. Each class represented a school district within a rural midwestern cooperative. The students' mean reading percentile rank was 45.1 (SD = 27.8) 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-Level 11 Mastery Test (MT; Clymer, Blanton, Johnson, & Lapp, 1980) of the Ginn

720 and Ginn 720 Rainbow Edition reading series were employed as measures. Each of the four scales, Comprehension, Vocabulary, Decoding, and Study Skills, 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 MT is criterion-referenced, with items per subtest ranging from 6 to 25 and with mastery-nonmastery cutoff scores established at 79% to 86% 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 MT. 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 MT, and five to six minutes for the Word Reading Test.

To assess test-retest reliability questions, a subgroup of 22 students (12 M, 10 F) was administered the following measures in the following order within a 2-week time period: the MT, the SRA Reading Achievement Test, the Word Reading Test, and the MT again. For the remaining 20 students, each measure was given one time within a 3 week period, with the order of administration random.

Data Analysis

Consistency of performance on two administrations of the same test. Consistency of students' performance on the MT was assessed in three ways. In all three analyses, the students who had been tested twice on the MT (N=22) were the subjects. First, traditional test-retest reliability was determined by correlating scores from the two

administrations of the MT. The other two analysis strategies were designed specifically for criterion-referenced measures (see Millman, 1974). In the first of these, consistency of students' subtest scores was determined by (a) computing individuals' percentage correct score on each subtest for each administration of the MT, (b) calculating for each individual his/her difference score across the two administrations of each subtest, and (c) determining the percentages of examinees having each possible difference score on each subtest. In the second strategy, consistency of mastery-nonmastery decisions on subtests 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 validity. The criterion validity of the MT was determined in two ways, employing the entire group of subjects (N=47). The traditional psychometric strategy of correlating scores on the measure of interest (MT) 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 mastery-nonmastery represented one dimension of each table and pre-post instructional status represented the other dimension. Percentages of misclassifications and phi coefficients supplemented the chi-square tests.

Results

Table 2 is a display of students' mean scores and standard deviations on each subtest of the MT, on the subscale 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 subtests of the MT are displayed in Table 3. For the comprehension subtests, correlations were moderate, ranging from .74 to .86; for the vocabulary and decoding subtests, correlations were high, ranging from .84 to .91; for the study skills subtests, correlations were low to moderate, ranging between .49 and .64. All coefficients for the test scales and total test were at least .90 with the exception of the Study Skills Scale, which had a coefficient of .69.

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 MT. Figures 1-2 are graphic displays of the percentages of examinees displaying various difference scores on each subtest of the MT; Table 4 summarizes the information illustrated on the graphs. The range of

difference scores on the subtests fell between 0 and 57%. The percentage of examinees with 0% difference scores on two administrations ranged from 18 on the inferential comprehension subtest to 73 on the prefixes subtest. Across the comprehension subtests, the mean percentage of examinees with 0 difference scores was 22.50 (SD = 6.36); across the vocabulary subtests; the mean percentage was 41.00 (SD = 19.79); across the decoding subtests, the mean percentage was 59.00 (SD = 19.79); across the study skills subtests, the mean percentage was 31.50 (SD = 6.36); and across all the subtests, the mean percentage was 38.50 (SD = 18.24).

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 MT. 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 comprehension, vocabulary, and decoding subtests, the corrected proportions were high, ranging from a proportion of agreement on the word meaning subtest of 55% higher than chance to a proportion of agreement on the prefixes subtest of 100% greater than chance. On the study skills subtests, the proportion of agreement was low for the respellings and accents subtest (29% greater than chance), but higher for the parts of an outline subtest (61% greater than chance).

Insert Table 5 about here

Criterion Validity

Correlational analyses were conducted between the MT subtests and two criterion measures, the SRA Reading Achievement Test and the Word Reading Test. Correlations between the MT subtests and the SRA Subscale and Total Test scores are displayed in Table 6. They ranged from .48 to .78 (SD=.09) when SRA vocabulary subscale scores were involved; from .36 to .73 (SD=.10) when SRA comprehension subscale scores were employed; and from .30 to .55 (SD=.07) when SRA Total Scores were used. The mean correlation for MT comprehension subtests was .55 (SD=.11); for the vocabulary subtests, .55 (SD=.12); for decoding subtests, .47 (SD=.13); and for study skills subtests, .50 (SD=.08).

Insert Table 6 about here

Correlations between the MT subtests and the Word Reading Test subscale scores are displayed in Table 7. They ranged from .31 to .82 when isolated word reading scores were involved, and from .33 to .85 when passage reading scores were employed. The mean correlation for the MT comprehension subtests was .56 (SD=.18); for the MT vocabulary subtests, .77 (SD=.08); for the MT decoding subtests, .56 (SD= .14); and for the MT study skills subtests, .44 (SD=.14).

Insert Table 7 about here

Criterion validity also was examined by inspecting the relation between mastery-nonmastery decisions on the MT and actual pre-post instructional status. Relevant chi-square values, phi coefficients, and percentages of misclassified students are displayed in Table 8. Although all chi-squares were significant, the level of misclassification and relationship between actual reading level and criterion performance showed more modest effects. Across the comprehension subtests of the MT, the average percentage of misclassified students was 15.0 (SD=9.90); across the vocabulary subtests, 20.0 (SD=6.36); across the decoding subtests, 37.0 (SD=6.36); across the study skills, 22.0 (SD=9.90); and across all the subtests, 21.0 (SD=7.23). The phi coefficient describing the relationship between reading level and criterion performance, was generally quite moderate, ranging from .29 to .77, with a median relationship of .68 for the subtest composites.

Insert Table 8 about here

Discussion

The purpose of the current study was to describe the reliability and criterion-related validity of a typical basal reading series, criterion-referenced mastery test. The study examined two aspects of the technical adequacy of the Ginn 720 End-of-level 11 Mastery 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 Ginn 720 MT total score seemed adequate. Some subtests appeared adequate; others did not.

Test-retest reliability coefficients indicated that, when the MT was administered twice within a short time interval, students' performance was somewhat inconsistent on the study skills subtests and on the literal comprehension subtest; none of the correlations obtained for these subtests even fell within the acceptable range for making group decisions (Salvia & Ysseldyke, 1981). Nevertheless, for the remaining comprehension subtest, for all of the vocabulary and decoding subtests, and for the total score and all scales except study skills, 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 with 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 80% better than chance agreement except for the respellings and accents, the parts of an outline, the word meaning, and the literal comprehension subtests; only the corrected proportion for the respellings and accents subtest fell below 55% better than chance.

Inspection of the consistency of test scores displayed in Figure 1 and 2 and in Table 4 reveals that the percentages of examinees

scoring the same across two administrations of the MT were variable. The average percentage of subjects obtaining the same score across all subtests was 38.5, with the percentages for the decoding and vocabulary subtests relatively high and with percentages for the comprehension and study skills subtests comparatively low. Given the fact that there are relatively few items in many subtests and given a mastery criterion of 79% to 86% per subtest, one might expect a difference of one or two items correct in an administration of an MT subtest to result in different mastery decisions. Interestingly, however, there appears to be no relation between the numbers of items in subtests and either the proportions of examinees placed into the same decision category on the two administrations or the test-retest reliability coefficients.

Therefore, although the results displayed in Figures 1 and 2 and in Table 4 appear variable and somewhat low for both comprehension subtests, the other consistency analyses generally support the adequacy of the Ginn 720 End-of-level 11 Mastery Test, with the exception of the study skills subtests. These findings are contrary to those of Tindal et al. (1983) who examined the End-of-level 11 Houghton-Mifflin Basic Reading Test and found that (a) the reliabilities of the study skills subtests were higher than those of the decoding and comprehension subtests, and (b) the results of the test-retest correlations and the corrected proportions of subjects placed into the same decision categories were lower than the results of the analysis involving the percentages of examinees who had different percentage correct scores across the two testings. Since,

in both studies, the analysis of the consistency of mastery/nonmastery decisions was less variable and agreed better with the test-retest correlations than did the analysis of the consistency of test scores, one might conclude tentatively that the consistency of decisions analysis is a more useful, and perhaps more valid, strategy for examining the reliability of criterion-referenced tests.

The criterion validity of the MT also was examined in this study. The traditional correlational analyses indicated that the criterion validity of the MT with respect to the SRA Reading Achievement Test was marginal, with only 15% of correlations between the MT and the SRA Scales falling above .70 and none at or above .80. Correlations between the MT and the Word Reading Test scales were generally higher, with 38% falling above .70 and 23% at or above .80. Given that both the Word Reading Test and the MT are curriculum-based, one might expect a stronger relation between these two measures than between the MT and the SRA Reading Achievement Test. Nevertheless, correlations among curriculum-based measures and more global indices have been reported frequently at high levels (Fuchs et al., in press; Fuchs, Fuchs, & Deno, 1982). This indicates that the correlations between the mastery and SRA tests are comparatively low, and that performance on the MT predicts concurrent performance on more global measures of reading proficiency relatively poorly.

The criterion validity of the MT also was investigated with the criterion-referenced strategy of examining the relation between the mastery-nonmastery classification on the MT and actual pre-post instructional status. Percentages of misclassifications ranged from

15% to 46% on the subtests, with 21% 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 MT was designed, the MT subtests have limited utility whereas the total test score is more valid.

Consequently, the current study suggests that the End-of-level 11 MT varied in quality. For predicting global reading proficiency, the usefulness of the MT appeared limited. However, for making decisions about student placement and progress within the curriculum, results were more favorable. Although in most analyses the study skills subtests appeared inadequate and the comprehension subtests were of variable quality, the decoding and vocabulary subtests were acceptable, and the total MT score was generally reliable and valid. This indicates that (a) educators should use the MT judiciously, relying primarily on the decoding subtests, vocabulary subtests, and total test scores for making decisions about mastery in the curriculum, and (b) test developers at Ginn and Co. might consider reexamining the study skills and comprehension subtests.

In any case, the technical adequacy of the End-of-level 11 Ginn MT was superior to that of the End-of-level 11 Houghton-Mifflin Basic Reading Test (Tindal et al., 1983). Interestingly, compared to Houghton-Mifflin (Wallis, 1983), Ginn and Co. (Walker, 1981) described preliminary examination of the quality of their mastery tests more thoroughly and appropriately. Perhaps, Ginn's somewhat more deliberate and empirical approach to test development at least partially explains the better reliability and validity coefficients of

their MT documented in this study. If so, this underscores two points: (a) content and face validity are necessary but insufficient dimensions of criterion-referenced test adequacy, and (b) test consumers must demand empirical validation of criterion-referenced tests before relying on such test data for making instructional decisions.

References

Clymer, T., Blanton, W. E., Johnson, D. D., & Lapp, D. Mastery test: Tell me how the sun rose, Reading 720, Level 11. Lexington, MA: Ginn and Co., 1980.

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., 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.

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.

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

Walker, R. Personal communication, January, 1983.

Wallis, H. Personal communication, January 13, 1983.

Table 1

Examinees' Tasks on the Ginn 720 End-of-Level 11 Mastery Test

Scale/Subtest	Examinees' Tasks
Comprehension	
Literal Comprehension	Read a factual article comprised of five paragraphs. Then, answer each of 10 questions by selecting the correct response from an array of three choices.
Inferential Comprehension	Read a selection. Then, answer a set of questions by selecting, for each question, a correct response from an array of three choices.
Vocabulary	
Word Meaning	Given a word or phrase, identify a synonym from an array of three choices.
Context Clues	Given a sentence with an underlined word, select a synonomous word or phrase from an array of three choices.
Decoding	
Prefixes	Given a sentence with one word omitted with a blank space, select a word that best completes the sentence from an array of three choices.
Suffixes	Given a sentence with one word omitted with a blank space, select a word that best completes the sentence from an array of three choices.
Study Skills	
Respellings and Accents	Read a sentence containing one underlined word. Given two respellings with pronunciations and accents for the underlined word, select the correct respelling.
Parts of an Outline	Read a four paragraph article. Then, given a partially completed outline, select, from an array of three choices, a word or phrase to complete correctly each omission from the outline.

Table 2

Student Performance on Measures of Reading Achievement (N=42)

Test	Mean	SD
End-of-level 11 Mastery Test		
Comprehension Subtests	24.6	6.0
Literal Comprehension	6.8	2.3
Inferential Comprehension	17.9	4.6
Vocabulary Subtests	25.2	7.6
Word Meaning	17.9	5.6
Context Clues	7.3	2.3
Decoding Subtests	23.7	6.2
Prefixes	4.8	1.7
Suffixes	18.9	4.8
Study Skills	9.1	2.2
Respellings and Accents	5.2	0.8
Parts of an Outline	3.9	1.8
Total Test	82.6	19.3
SRA Reading Achievement Test		
Vocabulary	28.2	7.8
Comprehension	30.9	12.2
Total	56.2	20.0
Word Reading Test		
Isolated Word Reading	50.1	24.7
Passage Reading	109.0	37.5

Table 3

Test-retest Reliabilities for Ginn 720 End-of-level 11 Mastery Test (N=22)

Subtest	Reliability
Comprehension Subtests	.93
Literal Comprehension	.74
Inferential Comprehension	.86
Vocabulary Subtests	.97
Word Meaning	.91
Context Clues	.90
Decoding Subtests	.90
Prefixes	.90
Suffixes	.84
Study Skills Subtests	.69
Respellings and Accents	.49
Parts of an Outline	.64
Total Test	.97

Table 4

Proportion of Subjects with Varying Percentages of Difference Scores Across Two Administrations of the End-of-level 11 Mastery Test (N=22)

Mastery Test	N[a]	Percentage Difference Score									
		0 to .07	.08 to .14	.15 to .24	.25 to .34	.35 to .44	.45 to .54	.55 to .64	.65 to .74	.75 to .84	.85 to 1.00
Comprehension Subtests											
Literal Comprehension	10	.27	.32	.23	.14	.05					
Inferential Comprehension	25	.50	.33	.10	.10						
Vocabulary Subtests											
Word Meaning	25	.59	.33	.10							
Context Clues	10	.55	.32	.14							
Decoding Subtests											
Prefixes	6	.73	0	.27							
Suffixes	24	.50	.42	.05	.05						
Study Skills											
Respellings and Accents	7	.36	.59	0	.05						
Parts of an Outline	7	.27	.41	0	.18	.05	0	.05			

[a]Number of items on the subtest.

Table 5

Uncorrected and Corrected Proportions of Examinees (N=22) Placed Into the Same Decision Categories on Two Administrations of the End-of-level 11 Mastery Test

Subtest	Proportion of Examinees	
	Uncorrected	Corrected for Chance Agreements[a]
Comprehension Subtests		
Literal Comprehension	.86	.71
Inferential Comprehension	.95	.91
Vocabulary Subtests		
Word Meaning	.77	.55
Context Clues	.91	.81
Decoding Subtests		
Prefixes	1.00	1.00
Suffixes	.95	.88
Study Skills Subtests		
Respellings and Accents	.68	.29
Parts of an Outline	.82	.61

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

Table 6

Correlations Between End-of-level 11 Mastery Test and SRA Test Scores (N=42)

Ginn Subtest	SRA Vocabulary	SRA Comprehension	SRA Total
Comprehension Subtests	.70	.73	.53
Literal Comprehension	.52	.58	.38
Inferential Comprehension	.65	.66	.50
Vocabulary Subtests	.70	.64	.47
Word Meaning	.73	.63	.48
Context Clues	.52	.57	.43
Decoding Subtests	.65	.51	.44
Prefixes	.48	.36	.30
Suffixes	.66	.52	.47
Study Skills Subtests	.69	.64	.48
Respellings and Accents	.54	.48	.41
Parts of an Outline	.60	.55	.39
Total Test	.78	.72	.55

Table 7

Correlations Between End-of-level 11 Mastery Test and Word Reading Test Scores (N=42)

Ginn Subtests	Word Reading Test	
	Isolated Words	Passages
Comprehension Subtests	.69	.72
Literal Comprehension	.37	.45
Inferential Comprehension	.71	.72
Vocabulary Subtests	.81	.85
Word Meaning	.82	.84
Context Clues	.67	.74
Decoding Subtests	.65	.65
Prefixes	.44	.44
Suffixes	.68	.69
Study Skills Subtests	.52	.58
Respellings and Accents	.31	.33
Parts of an Outline	.52	.59
Total Test	.80	.83

Table 8

Relation Between End-of-level 11 Mastery Test and Criterion Classification (N=46)

Subtest	χ^2	p	ϕ	Percentage Misclassified
Comprehension Subtests	24.36	.000	.73	15
Literal Comprehension	17.11	.002	.61	22
Inferential Comprehension	23.86	.000	.72	20
Vocabulary Subtests	23.90	.000	.72	20
Word Meaning	27.40	.000	.77	20
Context Clues	20.37	.000	.67	33
Decoding Subtests	18.85	.000	.64	37
Prefixes	12.83	.010	.29	46
Suffixes	21.97	.000	.53	35
Study Skills Subtests	13.04	.010	.53	22
Respellings and Accents	13.77	.010	.55	24
Parts of an Outline	9.81	.040	.46	26
Total Test	24.86	.000	.74	15
Median for Subtests	21.38	.000	.68	21

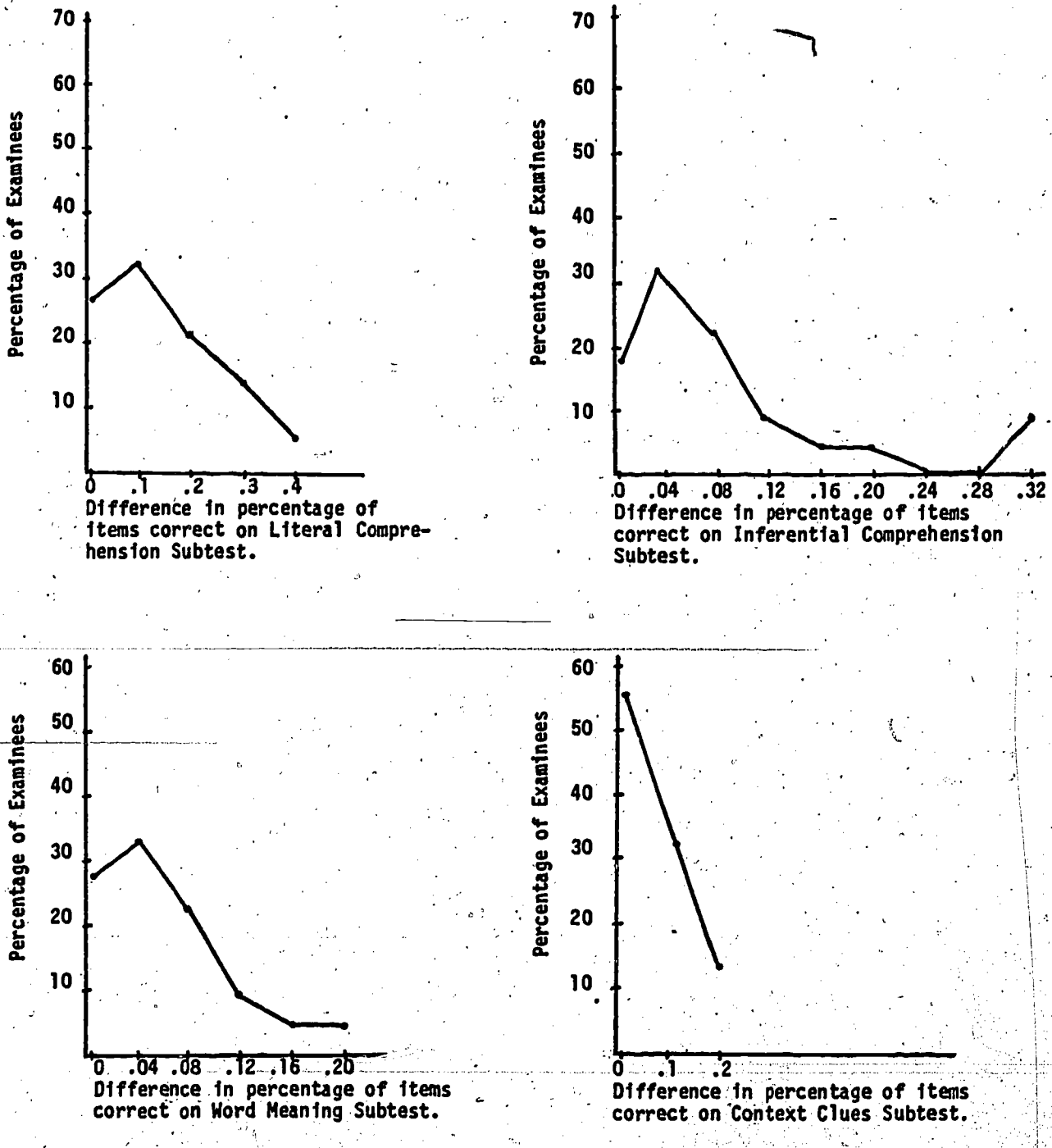
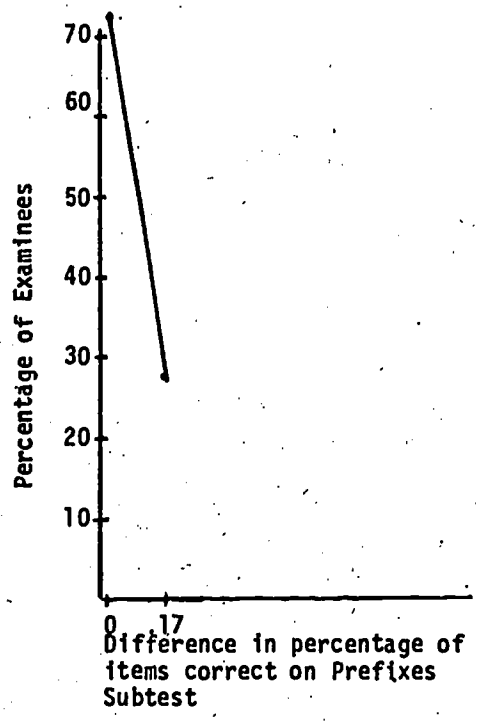
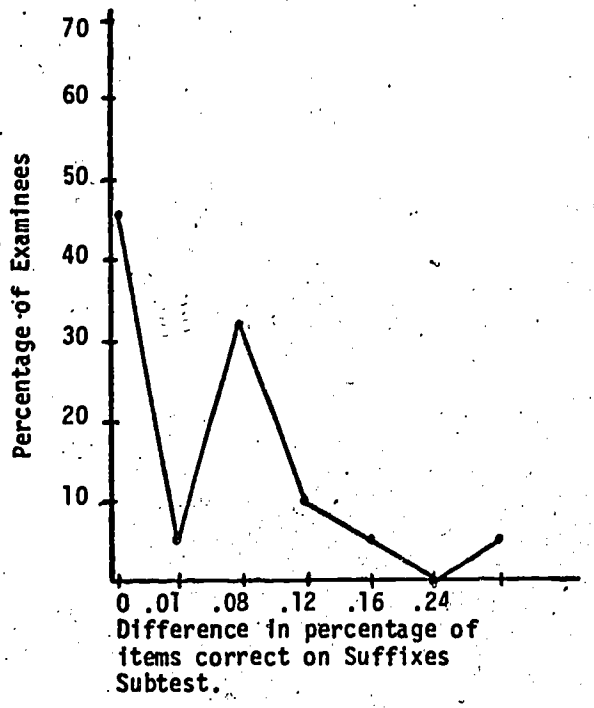
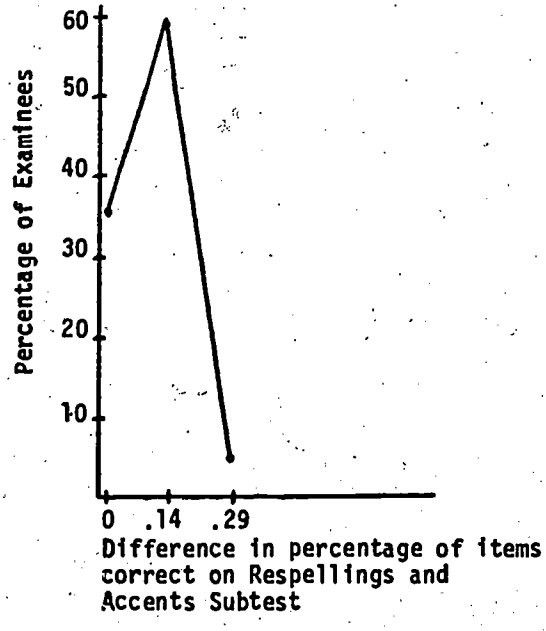


Figure 1. Displays of consistency of test scores on comprehension and vocuablary subtests of end-of-level 11 MT.







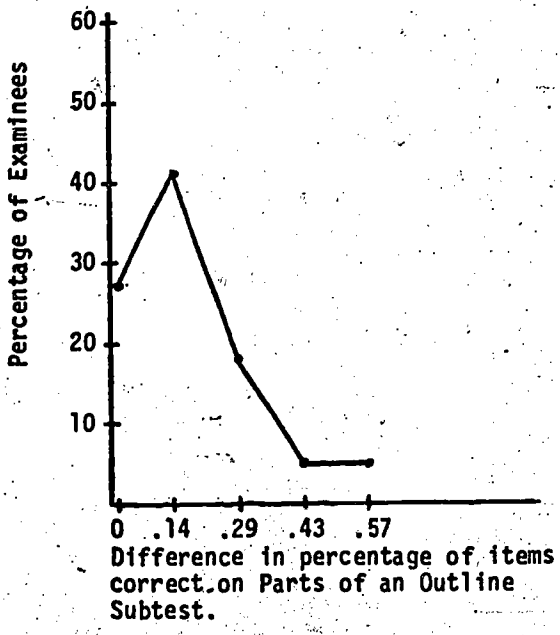


Figure 2. Displays of consistency of test scores on decoding and study skills subtests of end-of-level 11 MT.

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.