DOCUMENT RESUME

ED 236 198 TM 830 707

AUTHOR	Marston, Doug; And Others
TITLE	A Comparison of Standardized Achievement Tests and Direct Measurement Techniques in Measuring Pupil Progress.
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-126
PUB DATE	Jul 83
CONTRACT	200-80-0622
NOTE	29p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Achievement Gains; Elementary Education; Individual Differences; Language Tests; *Measurement Techniques; Pretests Posttests; *Reading Improvement; Reading Tests; Special Education; *Standardized Tests
IDENTIFIERS	*Direct Assessment; *Stanford Achievement Tests

ABSTRACT

The effectiveness of direct measurement techniques and standardized achievement tests for assessing within-individual change over a 10-week period was examined. The Reading Comprehension and Language subtests from the Stanford Achievement Tests and direct measures of reading and written language were administered twice to 83 low-achieving students in grades 3-6. Analyses indicated that greater student gains were evident on the direct measures than on the standardized achievement test. (Author)

University of Minnesota

Research Report No. 126

A COMPARISON OF STANDARDIZED ACHIEVEMENT TESTS AND DIRECT MEASUREMENT TECHNIQUES IN MEASURING PUPIL PROGRESS

Doug Marston, Stanley Deno, and Gerald Tindal

IRLD

Institute for Research on Learning Disabilities

IRLD

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).

Research Report No. 126

A COMPARISON OF STANDARDIZED ACHIEVEMENT TESTS AND DIRECT MEASUREMENT TECHNIQUES IN MEASURING PUPIL PROGRESS

Doug Marston, Stanley Deno, and Gerald Tindal

Institute for Research on Learning Disabilities

University of Minnesota

July, 1983

Abstract

The effectiveness of direct measurement techniques and standardized achievement tests for assessing within-individual change over a 10-week period was examined. The Reading Comprehension and Language subtests from the Stanford Achievement Tests and direct measures of reading and written language were administered twice to 83 low-achieving students in grades 3-6. Analyses indicated that greater student gains were evident on the direct measures than on the standardized achievement test.

A Comparison of Standardized Achievement Tests and Direct Measurement Techniques in Measuring Pupil Progress

How is the progress of exceptional children receiving special education services best measured? Traditionally, standardized achievement tests have been preferred for the assessment of student growth (Mehrens & Lehmann, 1973; Stanley & Hopkins, 1972). Student improvement, or lack of it, on such tests is perceived as an index of how the pupil is progressing in his or her school situation. On the basis of this information, much of what happens to a student served in special education is determined, including: program placement, program planning, and exit from special services. However, it can be argued that norm-referenced, standardized achievement tests do not effectively measure student learning (Carver, 1974).

The inadequacy of standardized achievement tests for measuring student change is related to three factors. First, norm-referenced achievement tests are designed primarily to measure individual differences, not changes in learning (Hively & Reynolds, 1975). Scores from these tests may be interpreted only in relation to the performance of others, and cannot be used for within-individual comparisons. Furthermore, Carver (1974) noted that a single test cannot fulfill both responsibilities. For instance, the most efficient item for a norm-referenced test is one that has a passing proportion of .50 (p = .50), which maximizes the population variance. However, an assessment procedure that best measures learning would have items with p values near .00 before educational intervention and approaching 1.00 after treatment. Thus, psychometrically sound, norm-referenced tests are not the optimal methodology for measuring individual student learning.

A second reason for dissatisfaction with standardized tests as measures of progress relates to the sensitivity of these devices in measuring what the student is taught. Jenkins and Pany (1978) demonstrated that standardized tests of reading achievement differentially sample the content of frequently used reading curricula. Examination of their hypothetical data should alert special educators to the distinct possibility that measurement of student growth is a function of the test used and does not necessarily reflect true changes in pupil performance. Lovitt and Eaton (1972) cited actual case data that corroborate this conclusion.

Third, the use of grade equivalent scores, a common practice with standardized achievement tests, is problematic in measuring student progress. It is well documented that grade equivalent scores are not expressions of equal interval units (Salvia & Ysseldyke, 1981; Thorndike & Hagen, 1969). Consequently, the aggregation or averaging of these scores in evaluating the progress of the special education pupil must be viewed as highly suspect.

Carver (1974) proposed that alternatives to norm-referenced psychometric methods must be developed if pupil progress is to be validly indexed. He labeled this new type of assessment as edumetric measurement, and emphasized the need for technically adequate procedures to evaluate within-individual gain. Recognizing the need for an edumetric approach, Jenkins, Deno, and Mirkin (1979) outlined the desirable characteristics of such a measurement system: it must be relevant to the child's curriculum, sensitive to growth, flexible and adaptive to various instructional objectives, repeatedly administrable, and easily administrable.

Several investigators in the field of special education have developed measurement systems focusing on the assessment of within-individual change (Deno & Mirkin, 1977; Lindsley, 1971; White & Haring, 1980). Recent research with these methods, often referred to as direct measurement techniques, has shown that it is possible to validly measure student behaviors in the classroom and satisfy most of the desired characteristics outlined by Jenkins et al. (1979). In the area of reading, Deno, Mirkin, and Chiang (1982) demonstrated that a student's oral reading rate on a passage from his or her basal reader or on a list of words from the reader correlated highly with standardized achievement tests of decoding (*r* = .90) and reading comprehension (*r* = .80). In a similar study, focusing on written language skills, Deno, Marston, and Mirkin (1982) examined the written compositions of normal and learning disabled elementary students and found that the number of words written, the number of words spelled correctly, and the number of correct letter sequences written all correlated highly with standardized tests of written language achievement (range = .70 - .86).

Such investigations are a necessary precondition in the establishment of direct measurement techniques as technically adequate (American Psychological Association, 1974). Another issue remains, however--the issue of the measures' sensitivity to growth. The capability of standardized achievement tests to monitor student change accurately has been seriously challenged (Hively & Reynolds, 1975). If direct measurement strategies are to be considered bona fide edumetric assessment procedures, their sensitivity to monitoring short-term progress must be substantiated.

It is the purpose of this paper to examine the effectiveness of direct measurement techniques to assess within-individual change in contrast to standardized achievement tests. Specifically, the study focuses upon the capacity of standardized achievement tests and direct measurement techniques to measure pupil progress in children with learning difficulties. A 10-week period was selected as the interval in which student progress in reading and written language would be assessed. Central to the analysis is the assumption that the measurement approach most sensitive to student growth would show the greatest pupil gains.

Method

Subjects

Low-achieving elementary students from grades 3-6 participated in the study. These students attended three schools located in a rural, midwestern area. A measure of written expression, validated by Deno, Marston, and Mirkin (1982), was used to select the low-achieving students from a total population of 785 pupils. All students were asked to write two compositions, with the total number of words written on the second composition tallied. Those students who performed at or below the 15th percentile for their grade level were invited to participate in the study. Parental permission was received for 83 pupils. The sample sizes, means, 15th percentile cutoff scores, and distribution of sexes for each grade level are presented in Table 1.

Insert Table 1 about here

Procedures

A set of standardized achievement tests and direct measurement tasks were administered twice to the 83 students. The assessment procedures were administered in the first week of October and 10 weeks later, in December.

The standardized achievement tests administered to the students were the Reading Comprehension and the Language subtests from the Stanford Achievement Tests (Madden, Gardner, Rudman, Karlsen, & Merwin, 1978). Because the grade levels of the students ranged from third to sixth grade, four different forms of the SAT were administered: Primary II-A, Primary III-B, Intermediate I-A, and Intermediate II-B. The scores obtained for each subtest were raw score, scaled score, grade equivalent, and percentile.

The direct measure of reading used in the study was derived from Deno, Mirkin, and Chiang (1982). A list of words was selected randomly from the third grade level of the Harris-Jacobson Word List (Harris & Jacobson, 1972) and used for the reading task. Each student was asked to read aloud for one minute. Test instructions read verbatim to the subject were:

> Here is a word list that I want you to read. When I tell you to start, you can read across the page. Please read as fast and accurately as you can. If you get stuck on any of the words, move on to the next one. I will tell you when to stop reading. Are there any questions? Ready? Begin.

The child then was timed for 60 seconds while the examiner followed along on a recording sheet identical to the student's list, recording the mistakes. If a student did not respond after approximately six seconds, he or she was told to move on to the next word. For each student the number of words read correctly (WRC) was scored. Estimates of inter-rater agreement ranged from .94 to .98.

In addition to reading the third grade lists, the fourth, fifth, and sixth graders were asked to read a list of words produced from their grade level from the Harris-Jacobson list. For example, the fifth graders read both a third grade list and a fifth grade list. For each student the number of words read correctly from grade level (WRCG) was counted.

The measure of written language employed in this study was based upon the research of Deno, Marston, and Mirkin (1982). Pupils were administered the same story starter at weeks 1 and 10; each time, they were given three minutes to complete this task. For each student, the total number of words written (TWW), the number of words written (spelled) correctly (WWC), and the number of correct letter sequences (CLS) were computed. Inter-scorer reliability coefficients ranged between .91 and .96.

Analysis

Two different analyses were conducted upon the data. Analysis I focused upon the amount of change from week 1 to week 10 on the standardized and direct measures. Using a paired *t* test analysis for each measure, a *t* value was computed and interpreted as representative of the amount of change that each test measured. While Analysis I provided an interpretation of how much growth is evident for each measure, it did not provide a direct comparison of standardized tests and direct measurement techniques. Analysis II was designed to produce this comparison by contrasting student change on direct measures with change on the standardized tests. However, assessment of improvement between weeks 1 and 10 on the measures was not made with equivalent units. Derived scores from the SAT included raw scores, scaled scores, grade equivalents, and percentiles. All scores on the direct measures were in raw score units. To remedy this situation, a modification of Glass' (1978) Effect Size analysis was used. Thus, for each student an Effect Size (ES) was calculated by subjecting the week 10 score to a type of z score transformation using week 1 as a referent.

In to the following formula, ES is a student's standardized week 10 score, X_2 is the observation at week 10, $\overline{X}_1$ is the mean of all students at week 1, and SD_1 is the standard deviation of all students at week 1:

$$ES = \frac{X_2 - \overline{X}_1}{SD_1}$$

Thus, the student's growth is determined by the ratio of his or her deviation from the week 1 mean and the standard deviation of week 1. If there is little or no change between weeks 1 and 10 on the measure, the student's ES approaches zero. However, if the student performed better at week 10, the ES should be greater than 0 (conversely, if week 10 scores were lower than week 1, ES would be less than 0). The transformation provides an index of student growth relative to initial performance that is directly comparable across measures. Once Effect Size scores were computed for each student on each assessment procedure, it was possible to compare student change performance on standardized tests and a direct measure. In Analysis II, the contrast of standardized and direct measures was achieved with paired t-test analysis with Effect Size as the dependent variable.

Results

Analysis I

Table 2 is a summary of the t values comparing the scores of weeks 1 and 10 for the standardized measures. For the Reading Comprehension subtest from the SAT, t values ranged from 2.17 to 2.55; all were significant ($p < .05$). Change was not apparent on the Language subtest from the SAT; t values comparing performance at weeks 1 and 10 ranged from .39 to 1.63, all statistically nonsignificant. As is evident in Table 3, t values for the direct measures were much larger; all were significant at the .001 level. The greatest change in performance was reflected in words read correctly (WRC: $t = 11.74$, $p < .001$). The smallest t value was found for words read correctly from grade level (WRCG: $t = 3.69$, $p < .001$).

Insert Tables 2 and 3 about here

Analysis II

Paired t test analyses comparing the mean student Effect Sizes for reading measures are presented in Table 4. For the area of reading, the student ESs for words read correctly (WRC) were significantly greater than all Reading Comprehension ESs ($p < .001$). Effect Sizes for words read correctly from grade level (WRCG) and SAT reading scores were not significantly different. In written language, total number of words written (TWW), number of words written correctly (WWC), and number of correct letter sequences (CLS) student Effect Sizes were all significantly greater than the SAT Language Effect Sizes ($p < .001$). These values are presented in Table 5.

Insert Tables 4 and 5 about here

Discussion

The measurement of pupil progress is a significant issue for those responsible for the delivery of special education services. PL 94-142 mandates that an Individual Educational Plan (IEP) be written for each handicapped child; the IEP includes the specification of goals and objectives related to the pupil's instructional needs. As Jenkins et al. (1979) pointed out, the implementation of such a system should "raise our sensitivities about the need to develop satisfactory procedures for measuring children's progress" (p. 82). Yet, debate continues over the appropriateness of standardized achievement measurement as the primary methodology for monitoring a student's progress over brief time intervals.

Direct measurement techniques are perceived by many as a viable alternative. Although the study of these techniques has been initiated only recently, it appears that the measures are valid with respect to APA Standards (Deno, Mirkin, & Chiang, 1982; Deno, Marston, & Mirkin, 1982). The analyses presented here provide preliminary evidence that direct measurement techniques are more sensitive to short-term growth in pupils with learning difficulties than are standardized achievement tests. In reading, the student Effect Sizes for oral fluency were significantly greater than the gains students made on the SAT Reading Comprehension subtest. Similarly, direct measures of written expression were much more sensitive to pupil progress over 10 weeks than the SAT Language subtest, on which virtually no growth was evident.

While the Effect Size analysis dramatically supports the notion that direct measurement is more sensitive to growth, the conclusion must be tempered by the absence of an external criterion for student improvement. It is plausible that student growth did not occur over the 10-week period, and that standardized achievement tests more accurately detected this phenomenon. However, this argument is based on the notion that students improved very little over a 2½ month period, an event that seems improbable, even for low-achieving pupils. Regardless, future research in this area must attend to this methodological deficiency.

Of greater concern may be the use of a pre-post test design to study pupil change in performance. The reliability of such change scores has been debated (Cronbach & Furby, 1970). In practice, this criticism would have little effect on direct measurement, for in addition to being a measurement system more closely linked to the student's curriculum, it also is a system based upon repeated measurements and not the pre-post test design. As Nunnally (1967) noted, repeated observations increase reliability. Standardized achievement tests, on the other hand, can only be used in pre-post test designs since they are not designed to be used on a repeated and frequent basis.

In summary, the needs of special educators in measuring pupil progress for IEP goals and objectives may be fulfilled by the use of direct measurement techniques. The preliminary research on pupil progress measurement presented here supports the contention that direct measures are preferable to standardized achievement tests.

References

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. Standards for educational and psychological tests. Washington, D.C.: American Psychological Association, 1974.

Carver, R. P. Two dimensions of tests: Psychometric and edumetric. American Psychologist, 1974, 29, 512-518.

Cronbach, L. J., & Furby, L. How we should measure "change" - or should we? Psychological Bulletin, 1970, 1, 68-80.

Deno, S. L., Marston, D., & Mirkin, P. K. Valid measurement procedures for continuous evaluation of written expression. Exceptional Children, 1982, 48, 368-371.

Deno, S. L., & Mirkin, P. K. Data-based program modification: A manual. Reston, VA: The Council for Exceptional Children, 1977.

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

Glass, G. V. Integrating findings: The meta-analysis of research. Review of Research in Education, 1978, 5, 351-379.

Harris, A. J., & Jacobson, M. D. Basic elementary reading vocabularies. New York: MacMillan, 1972.

Hively, W., & Reynolds, M. C. (Eds.). Domain-referenced testing in special education. Reston, VA: The Council for Exceptional Children, 1975.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. Measuring pupil progress toward the least restrictive alternative. Learning Disability Quarterly, 1979, 2, 81-92.

Jenkins, J., & Pany, D. Standardized achievement tests: How useful for special education? Exceptional Children, 1978, 44, 448-453.

Lindsley, O. R. Precision teaching in perspective: An interview with Ogden R. Lindsley. Teaching Exceptional Children, 1971, 3(3), 114-119.

Lovitt, T., & Eaton, M. Achievement tests vs. direct and daily measurement. In G. Senf (Ed.), Behavior analysis in education. Lawrence, KS: University of Kansas, 1972.

Madden, R., Gardner, E., Rudman, H., Karlsen, B., & Merwin, J. Stanford achievement test. New York: Harcourt Brace Jovanovich, 1978.

Mehrens, W. A., & Lehmann, I. J. Measurement and evaluation in education and psychology. New York: Holt, Rinehart & Winston, 1973.

Nunnally, J. Psychometric theory. New York: McGraw-Hill, 1967.

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

Stanley, J. C., & Hopkins, K. D. Educational and psychological measurement and evaluation. Englewood Cliffs, NJ: Prentice-Hall, 1972.

Thorndike, R. L., & Hagen, E. Measurement and evaluation in psychology and education. New York: John Wiley, 1969.

White, O. R., & Haring, N. G. Exceptional teaching (2nd ed.). Columbus, OH: Charles E. Merrill, 1980.

Table 1

Descriptive Data from Subject Selection Procedure

Grade	Number of Students Screened	Mean	15th Percentile Cutoff Score	Sex Distribution Male	Sex Distribution Female
3	190	17.7	9.0	14	12
4	185	22.7	12.0	10	7
5	225	30.5	19.0	14	5
6	185	36.4	24.0	13	8

Table 2

Paired t-test Comparison of Student Performance at Weeks 1 and 1

on Reading Comprehension and Language Subtests of the SAT

	Time of Assessment	Mean	Standard Deviation	t val
Reading Comprehension Raw Score	Week 1	41.42	21.7	2.3
	Week 10	43.74	21.3	
Reading Comprehension Scaled Score	Week 1	152.56	19.2	2.5
	Week 10	155.92	17.0	
Reading Comprehension Grade Equivalents	Week 1	5.05	1.9	2.1
	Week 10	5.31	1.8	
Reading Comprehension Percentile Ranks	Week 1	52.08	26.5	2.1
	Week 10	56.19	26.0	
Language Raw Score	Week 1	35.87	12.9	-1.0
	Week 10	34.40	12.2	
Language Scaled Score	Week 1	149.47	25.2	-.3
	Week 10	148.56	22.0	
Language Grade Equivalents	Week 1	4.77	2.1	-.5
	Week 10	4.66	2.0	
Language Percentile Ranks	Week 1	50.26	27.0	-1.6
	Week 10	45.19	26.8	

Table 3

Paired t-test Comparison of Student Performance at Weeks 1 and 10 on Direct Measures of Reading and Written Expression

	Time of Assessment	Mean	Standard Deviation	t value	Probability
Words Read Correctly 3rd Grade Level	Week 1 Week 10	46.85 60.71	26.1 28.8	11.74	.001
Words Read Correctly Grade Level	Week 1 Week 10	30.33 37.00	20.1 25.3	3.69	.001
Total Words Written	Week 1 Week 10	26.23 34.65	9.1 12.2	6.28	.001
Words Written Correctly	Week 1 Week 10	23.65 31.01	9.0 11.8	5.86	.001
Correct Letter Sequences for Writing Task	Week 1 Week 10	110.96 146.40	41.6 54.7	6.37	.001

Table 4

Comparison of Student Effect Sizes on SAT Subtests and Direct Measures in Reading*

Measures Compared[a]	Mean Student Effect Size	Standard Deviation	t value	Probability
DM: Words Read Correctly (3rd Grade)	.68	1.1	3.73	.001
ST: Reading Comprehension (Raw Score)	.13	1.2		
DM: Words Read Correctly (3rd Grade)	.70	1.1	4.16	.001
ST: Reading Comprehension (Scaled Score)	.16	1.0		
DM: Words Read Correctly (3rd Grade)	.70	1.1	4.20	.001
ST: Reading Comprehension (Grade Equivalent)	.15	1.0		
DM: Words Read Correctly (3rd Grade)	.70	1.1	3.98	.001
ST: Reading Comprehension (Percentile)	.18	1.0		
DM: Words Read Correctly (Grade Level)	.32	1.2	1.02	.313
ST: Reading Comprehension (Raw Score)	.15	1.0		
DM: Words Read Correctly (Grade Level)	.29	1.2	.57	.575
ST: Reading Comprehension (Scaled Score)	.21	.8		
DM: Words Read Correctly (Grade Level)	.29	1.2	.82	.415
ST: Reading Comprehension (Grade Equivalent)	.17	.9		
DM: Words Read Correctly (Grade Level)	.29	1.2	.80	.412
ST: Reading Comprehension (Percentile)	.18	1.0		

[a]DM represents direct measurement
ST represents standardized achievement test

Table 5

Comparison of Student Effect Sizes on SAT Subtest and Direct Measures in Written Language*

Measures Compared	Mean Student Effect Size	Standard Deviation	t value	Probability
DM: Total Words Written ST: Language (Raw Score)	1.26 -.19	1.7 1.1	4.98	.001
DM: Total Words Written ST: Language (Scaled Score)	1.26 -.10	1.7 1.0	4.69	.001
DM: Total Words Written ST: Language (Grade Equivalent)	1.26 -.08	1.7 1.1	4.62	.001
DM: Total Words Written ST: Language (Percentile)	1.26 -.38	1.7 1.2	5.63	.001
DM: Words Written Correctly ST: Language (Raw Score)	.94 -.19	1.6 1.1	4.13	.001
DM: Words Written Correctly ST: Language (Scaled Score)	.94 -.10	1.6 1.0	3.84	.001
DM: Words Written Correctly ST: Language (Grade Equivalent)	.94 -.08	1.6 1.1	3.78	.001
DM: Words Written Correctly ST: Language (Percentile)	.94 -.38	1.6 1.2	4.80	.001
DM: Correct Letter Sequences ST: Language (Raw Score)	1.15 -.19	1.8 1.1	4.40	.001
DM: Correct Letter Sequences ST: Language (Scaled Score)	1.15 -.10	1.8 1.0	4.10	.001
DM: Correct Letter Sequences) ST: Language (Grade Equivalent)	1.15 -.08	1.8 1.1	4.02	.001
DM: Correct Letter Sequences) ST: Language (Percentile)	1.15 -.38	1.8 1.2	4.96	.001

*DM represents direct measurement
ST represents standardized achievement test

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 learn[illegible] bility 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 pretest 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.