Reading Horizons: A Journal of Literacy and Language Arts

Volume 18
Issue 1 *October 1977*

Article 7

10-1-1977

Profile Analysis of Diagnostic Reading Tests: Is It Warranted?

Kenneth Kavale
University of Colorado, Boulder

Alfred Hirshoren
University of Georgia

PROFILE ANALYSIS OF DIAGNOSTIC READING TESTS: IS IT WARRANTED?

Kenneth Kavale
UNIVERSITY OF COLORADO, BOULDER, COLORADO
Alfred Hirshoren
UNIVERSITY OF GEORGIA

Through the use of various screening instruments and achievement measures, a number of children are identified as possible reading disability cases. Once identified, these children are subjected to further testing usually with a reading diagnostic test which attempts to identify specific strengths and weaknesses in performance. Discussions with school psychologists, educational diagnosticians, and teachers of children with reading disabilities suggest that much credence is still placed on profile analysis of diagnostic reading test data as suggested by their authors (Durrell, 1957: Karlsen, Madden, and Gardner, 1966) among others. Reviews by Wechsler (1974), Sattler (1974) and Hirshoren and Kavale (1977), in the area of intelligence testing, provide appropriate cautions regarding the practice of profile analysis. Since it is not unusual to find remedial programs developed at least in part on these profiles, it might be profitable to examine why profile analysis adds spurious specificity and misatriculated authority to quasi diagnostic statements.

Using the *Stanford Diagnostic Reading Test* (SDRT) Level II, form W as an example,[1] the pitfalls of profile analysis can be examined. When dealing with any measurement instrument, an important consideration is its stability and consistency which is expressed by the reliability coefficient. While the reliabilities of individual subtests in the SDRT are acceptable (r's range from .72 to .96), it is apparent that some subtests are less stable than others in measuring whatever the subtests do in fact measure. Reliability as an index of reproduce-ability is an important consideration in interpreting scores earned on separate subtests since the lower the reliability of a particular subtest, the less confidence there is that an accurate assessment of true ability has been achieved.

While the reliability coefficient expresses stability in a relative sense, a more useful statistic is the standard error of measurement (SEm) which expresses test reliability in an absolute sense, that is, with score units. The reliability coefficient and standard deviation provide an estimate of the standard error of measurement (SEm = SD $\sqrt{1-r}$) (Doppelt, 1956). This statistic indicates how much an individual's score would be expected to vary upon repeated examination with the same test.

---

[1] The SDRT is used only as a representative example of a type of test format and we do not mean to single it out for specific criticism.

Specifically, the standard error of measurement indicates the band of error surrounding a test score. It is used to define limits around an obtained score written which there is reasonable assurance that an individual's true score exists. At grade 5, the standard errors of measurement for the SDRT subtests range from 1.4 to 4.6 (p. 29). The reasonable limits around an obtained score can be determined by the level of confidence desired. If the usual .05 significance level is used, the 95% confidence interval can be calculated by multiplying the standard error of measurement of each subtest by 1.96 (± 1.96 SEm) (Hays, 1973). The resulting values indicate a 95 percent confidence that the individual's true score is not more than ± 1.96 SEm away from the obtained score. For example, in defining the 95% confidence interval at grade 5 for the SDRT, variations in subtest scores range from 6 to 18 raw score points around the obtained scores depending upon the particular subtest. On the SDRT, raw scores are converted to stanines on which the profiles are based. Figure 1 presents a hypothetical protocol in which the 95% confidence limits are portrayed for raw scores converted to stanines. (See figure one.)

The band portrayed in Figure 1 represents the fact that a score will be in this range 95% of the time. The most striking feature is the possible variation in scores for individual subtests. If stanines 4, 5, and 6 are considered average, it can be observed that the variations can range from average to superior (stanine 9), from average to below average (stanine 2 and 3), or average to poor (stanine 1). Thus, there is considerable variation in the amount by which a single obtained score may differ from the hypothetical true score due to errors of measurement. This would suggest that the interpretation of performance on a particular subtest is not warranted because of the considerable variation possible. Figure 2 represents the profile from the obtained raw scores converted into stanine equivalents. (See figure two.)

The resulting profile is then analyzed in an attempt to assess those specific reading skills which can be considered strong, adequate, or in need of remedial instruction. This approach, although attractive on the surface, may lead to the development of programs which actually work against the child's real needs. The reason is found in the magnitude of error present on subtest scores which allow for any number of theoretically possible profiles based upon the same hypothetical obtained scores (presented in Figure 2) and also within the 95% confidence interval. Figures 3 and 4 present other theoretically possible profiles based upon that same data. The result is a dilemma for the reading teacher: What is the true picture of the child's abilities? What remedial program will be most effective for this child? (See figures 3 and 4.)

The purported purposes of profile analysis are the identification of relatively strong and weak areas of functioning but an examination of Figures 2, 3, and 4 suggest that this goal is difficult to achieve because of lack of stability in individual subtests. The results are three different profiles which can be interpreted quite differently even though the total raw scores remain the same. While total performance is less resistant to variation because of high overall reliability, the larger error inherent in

FIGURE 1

95% CONFIDENCE BAND
FOR
SCORES CONVERTED TO STANINES

TEST	TEST 1: Reading Comprehension			TEST 2	TEST 3	TEST 4	TEST 5	TEST 6
	Literal	Inferential	Total	Vocabulary*	Syllabication	Sound Discrimination	Blending	Rate of Reading
RAW SCORE	20	15	35	33	13	28	8	11
STANINE	9	9	9	9	9	9	9	9
	8	8	8	8	8	8	8	8
	7	7	7	7	7	7	7	7
	6	6	6	6	6	6	6	6
	5	5	5	5	5	5	5	5
	4	4	4	4	4	4	4	4
	3	3	3	3	3	3	3	3
	2	2	2	2	2	2	2	2
	1	1	1	1	1	1	1	1

*Insert Reading Comprehension Total Grade Score here.

FIGURE 2

PROFILE BASED UPON RAW SCORES CONVERTED TO STANINES

TEST	TEST 1: Reading Comprehension			TEST 2	TEST 3	TEST 4	TEST 5	TEST 6
	Literal	Inferential	Total	Vocabulary	Syllabication	Sound Discrimination	Blending	Rate of Reading
RAW SCORE	20	15	35 *	33	13	28	8	11
STANINE	9	9	9	9	9	9	9	9
	8	8	8	(8)	8	8	8	8
	7	7	7	7	7	7	7	7
	6	6	(6)	6	6	6	6	6
	5	5	5	5	5	(5)	5	5
	4	4	4	4	(4)	4	4	4
	3	3	3	3	3	3	(3)	3
	2	2	2	2	2	2	2	(2)
	1	1	1	1	1	1	1	1

*Insert Reading Comprehension Total Grade Score here.

FIGURE 3

VARIATION I

T E S T	TEST 1: Reading Comprehension			TEST 2	TEST 3	TEST 4	TEST 5	TEST6
	Literal	Inferential	Total	Vocabulary	Syllabication	Sound Discrimination	Blending	Rate of Reading
RAW SCORE	20	15	35 [*]	33	13	28	8	11
	9	9	9	9	9	9	9	9
S	8	8	8	8	8	8	8	8
T	7	7	(7)	7	7	7	7	7
A	6	6	6	(6)	6	6	6	6
N	5	5	5	5	(5)	5	5	5
I	4	4	4	4	4	(4)	4	4
N	3	3	3	3	3	3	3	3
E	2	2	2	2	2	2	2	2
	1	1	1	1	1	1	(1)	(1)

*Insert Reading Comprehension Total Grade Score here.

FIGURE 4

VARIATION 2

TEST	TEST 1: Reading Comprehension			TEST 2	TEST 3	TEST 4	TEST 5	TEST 6
	Literal	Inferential	Total	Vocabulary	Syllabication	Sound Discrimination	Blending	Rate of Reading
RAW SCORE	20	15	35*	33	13	28	8	11
STANINE	9	9	9	(9)	9	9	9	9
	8	8	8	8	8	8	8	8
	7	7	7	7	7	(7)	7	7
	6	6	6	6	6	6	6	6
	5	5	5	5	5	5	5	(5)
	4	4	(4)	4	4	4	(4)	4
	3	3	3	3	3	3	3	3
	2	2	2	2	(2)	2	2	2
	1	1	1	1	1	1	1	1

*Insert Reading Comprehension Total Grade Score here.

individual subtests should preclude the use of profile analysis for the development of remedial programs.

Goodman (1968) and Smith (1971), in discussions of the reading process, emphasize that reading should be viewed as a global entity rather than a series of independent and uniquely defined skills. They further suggest that definitions of reading should avoid singling out any ability as crucial or overwhelmingly important. Thus, while profile analysis appears a logical and rational basis for educational planning, the fact that individual subtests representing single abilities are the foundation suggests caution in interpretation. Rather than being viewed as singular analyzable entities, the results of diagnostic tests should be interpreted in a holistic manner since subtests are only a means of assessing discrete abilities and do not highlight the complex interrelationships of those abilities to the total process.

The difficulties inherent in such tests as the SDRT, Durrell Analysis of Reading Difficulty (Durrell, 1957), or the Diagnostic Reading Scale (Spache, 1963) would suggest that they are best utilized in comparing a child's performance with a norm group. This use of diagnostic reading tests needs to be verified by the use of specific skill measures (Carver, 1972). The results may otherwise lead to the development of highly questionable remedial programs.

REFERENCES

Carver, R. P. Reading tests in 1970 versus 1980: Psychometric versus edumetric. *The Reading Teacher*, 1972, *26*, 229-302.

Doppelt, J. E. How accurate is a test score? *Test Service Bulletin,* 1956, *50*, 1-3.

Durrell, D. D. *Durrell Analysis of Reading Difficulty*. New York: Harcourt, Brace and Jovanovich, Inc., 1957.

Goodman, K. S. *The Psycholinguistic Nature of the Reading Process.* Detroit: Wayne State University Press, 1968.

Hays, W. L. *Statistics for the Social Sciences*, 2nd ed. New York: Holt, Rinehart, and Winston, Inc., 1973.

Hirshoren, A. and Kavale, K. Profile analysis of the WISC-R: A continuing malpractice. *The Exceptional Child,* 1976, *23*, 83-87.

Karlsen, B., Madden, R., and Gardner, E. F. *Stanford Diagnostic Reading Test,* Level II, Form W. New York: Harcourt, Brace and Jovanovich, Inc., 1966.

Sattler, J. M. *Assessment of Children's Intelligence* (rev. ed.). Philadelphia: Saunders, 1974.

Smith, F. *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read.* New York: Holt, Rinehart, and Winston, 1971.

Spache, G. D. *Diagnostic Reading Scales.* Monterey, California: California Test Bureau/McGraw-Hill, Inc., 1972.

Thorndike, R. L. and Hagen, E. *Measurement and Evaluation in Psychology and Education*, (3rd ed.). New York: John Wiley and Sons, 1969.

Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children* (rev. ed.). New York: Psychological Corporation, 1974.