DOCUMENT RESUME

ED 155 589 CS 004 079

AUTHOR Powell, William R.
TITLE Measuring Reading Performance Informally.
PUB DATE May 78
NOTE 17p.; Paper presented at the Annual Meeting of the International Reading Association (23rd, Houston, Texas, May 1-5, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Classroom Observation Techniques; Elementary Education; *Evaluation Criteria; *Informal Assessment; *Informal Reading Inventory; Oral Reading; *Reading Ability; Reading Comprehension; *Reading Diagnosis; Student Behavior; Word Recognition

ABSTRACT

To improve the accuracy of the informal reading inventory (IRI), a differential set of criteria is necessary for both word recognition and comprehension scores for different levels and reading conditions. In initial evaluation, word recognition scores should reflect only errors of insertions, omissions, mispronunciations, substitutions, unkown words, and transpositions; symptomatic behavior should not be considered. After the student has read, comprehension questions should be asked on a literal level, on implicit understanding, on vocabulary, and on evaluative skill; all should be wholly context dependent. Baseline criteria for determining unsatisfactory reading should be established in comprehension, word recognition, and symptomatic behavior; comprehension is the most significant in determining placement. The initial task of the IRI is to place the student at his reading level; this decision is first made on the basis of quantitative data, then an error analysis gives qualitative information. Placement precedes analysis, but both are necessary for effective reading diagnosis and placement. (DF)

William R. Powell
University of Florida
Gainesville, Fl. 32611
IRA Convention
Pre-Convention Institute on
Perspectives on Testing, #11
Houston, 1978

MEASURING READING PERFORMANCE INFORMALLY

William R. Powell
University of Florida

In the very early days of motion pictures when Western thrillers were shown in the ranch towns of the West, real cowboys would ride in from miles around to see the Saturday night show. As they sat in dingy theaters, they would watch the villain ride across the screen shooting up the town, robbing banks, and holding lovely heroines at his mercy. It is said that the cowboys sometimes became so excited and enraged at the villain that they would pull out their own guns and shoot at the screen.

The absurdity of this effort is readily apparent. If they had wanted to get the villain they would have done better to turn in their seats and shoot out the projector. The screen was only the reflection of the problem. The problem was somewhere else, difficult to get to, but no less real.

Reading professionals may be spending their time shooting at screens. At a reading conference in Chicago, Johns (1976) asked 24 professionals their views regarding 24 given statements about the informal reading inventory. It is interesting that those reading people were most concerned that the IRI overestimated the proper reading level for instruction (18 of 24 respondents). However, statements showing concern about the validity, reliability, the criteria for determining levels were ranked 15th, 16th, and 17th respectively. There was apparently little awareness on the part of these professionals that the problem for which they were most concerned was

a mere reflection of the things they rated so relatively insignificant.

The informal reading inventory (IRI) is a teacher constructed test which can be used to measure oral reading performance. It is a record of a reader's performance in a succession of short (125-175 words) reading passages of increasing difficulty. Steuver (1967) has shown that it takes a passage length in this range for the quantitative errors to stablize so a subsequent qualitative analysis may be done, if desired.

The record is obtained by having a student read each passage while the teacher monitors the oral reading performance and asks previously prepared, varied questions about each passage. The teacher keeps a record of the oral reading errors and responses to the questions. The reading passages should be taken from or should be equivalent in difficulty to books used for instruction in the classroom.

The purpose of an IRI is to allow a teacher to make a semi-controlled observation of the many characteristics of a child's reading performance. The results of an IRI should predict the book level, readability level, or level of difficulty a subject can process at three levels of performance according to specified criteria. It is certain characteristics of observed reading behavior and the specified criteria which are the focus of this paper.

Characteristics of Oral Reading Behavior

Primary and secondary analysis of oral reading data reveal that children's actual performances produce a reducing variable error pattern across levels. Easier material permits greater production of and latitude

for error than in increasingly more difficult material. There is an observed pattern of error ratios from high error ratios to low ones, from production of many errors to a reduced number of miscues. (Powell, 1970; Powell and Dunkeld, 1971; Dunkeld, 1970; Page, 1975; Spache, 1963; Durrell, 1955; Gilmore, 1968; Gray, 1963; Isakson and Miller, 1978.)

This reducing error pattern produces a differential effect of observed accuracy (error values) across levels. Performance in first grade material is not the same as a performance in sixth grade material. Error values of satisfactory performance is not the same for each level. There is an average of about a two-and-one-half percent difference between levels as the material increases in difficulty. This pattern creates an inverse relationship between levels of difficulty and latitude for error in acceptable ranges of performance. (Powell, 1970; Dunkeld, 1970)

The key to this movement across different reading levels is the difficulty or the readability of the material. Difficulty or maturity of written language, not the grade placement or age of the reader, is the primary factor effecting error production. (Powell, 1971; Watkins, 1953.) Kinsbourne (1976) has observed that when a youngster experiences difficulty in reading, the nature of his errors represents the level of difficulty of the material (expressed in readability grade level estimates) at which he is functioning rather than his chronological grade placement, chronological age or any relevant pathology. (p. 154)

The easier the material the more error that can be tolerated and still maintain satisfactory understanding. As the material increases in difficulty, the less error latitude is permitted for understanding. Students in material

pre-primer to second grade difficulty can cope with more error and maintain understanding than in material from the three to five range, which in turn can tolerate more error than in material of sixth grade difficulty or higher. Readers tolerate error in an inverse relationship to the difficulty of the material. (Powell, 1971; Page, 1975; Dunkeld, 1970.)

Support for these findings also comes from growth in cognitive development studies. Wohlwill (1962) indicates that the movement from perception to reasoning permits a decrease in dependence of behavior on information from the immediate stimulus field. This movement is due to a decrease in redundancy, an increase in selectivity, and an increase in the temporal and spatial contiguity tolerated. (pp. 98-102)

When a student reaches approximately the fifth grade instructional reading level, it is doubtful if the criterion of word recognition accuracy is a valid indicator for the differentiation. This finding was apparent in Dunkeld's data (1970), and Matejcek (1976) found a plateau effect at an earlier grade level in the phonetically consistent Czech language. Once a reader masters decoding skills to a level of automaticity, then that factor begins to lose its differential function. As a reader increases in reading power, the number of miscues gradually loses its strength as an indicator for differential value.

However, the number of recognition errors represent a relationship to the sample size of the subject's recognition vocabulary. If the number of words based on the word frequency studies of Carroll, Davies, and Richman (1971) is interacted with the word recognition criteria suggested by Powell

(1973), then it is possible to calculate the cumulative word corpus for each grade level data available. If the calculation is made at the point where frsutration level would begin, then the number of words that should form the base sight vocabulary for a given grade level can be determined. For example, at third grade level, a decrease in the error ratio results in an increase in sight vocabulary by about 275 words. Therefore, one error does make a difference as each additional error would reflect a decrease of sight vocabulary of approximately 250 to 300 words.

The types of miscues that are counted as errors in computing the word recognition ratio makes a difference, too. The evidence from Dunkeld's study (1970) would indicate that only insertions, omissions, mispronunciations, substitutions, unknown words, and transpositions should be counted as scorable errors. The inclusion of repetitions, self-corrections, hesitations, or any other type of symptomatic behavior does not enhance the accuracy of word recognition score. However, as Clay (1972) has so clearly indicated, the number of self-corrections to the total number of errors are of diagnostic value in revealing the efficiency of processing clues. This diagnostic insight is additional to the word recognition score and calculated separately from it.

In evaluating oral reading behavior via an informal reading inventory, a cautionary note needs to be applied to the comprehension score on the first passage read. It is likely to be an unreliable score, as observation (Dunkeld, 1970) has recorded that this is true in two-thirds of the protocols studies. Furthermore, it has been observed that symptoms (repetitions,

hesitations, finger point, word-by-word reading, holding book too close, etc.) are of diagnostic value only during the first two years of the learning to read process. (Dunkeld, 1970.) After that point, the temporarily induced behavior fades into habituation and loses its diagnostic significance.

Comprehension scores on an informal reading inventory has a differential criteria in proportion to the reducing effect of word recognition. Studies have shown a strong positive relationship between word recognition and comprehension. Ehri (1978) finds children must "read words not only accurately but also rapidly and with attention." (p. 18) Pace and Golinkoff (1976) state that although some reading theorists argue that accurate word recognition is not even necessary for the comprehension of the text, they conclude from their data that if a word cannot be decoded readily, its meaning may not be retrieved and access of meaning will not occur. (p. 765) "Comprehension may be impaired if too many words cannot be decoded readily." (p. 766) Reading comprehension deficiency is at least partially attributed to a failure to fully identify a sufficient number of words during the course of reading.

The data given in studies by Markman (1977) and Pace (1978) clearly show a differential level of comprehension accuracy across grade levels. Children in the lower grades have a less effective performance, due to passive processing, than those older youngsters who have to monitor and transform linguistic input. Assuming a baseline level can be established for comprehension sufficiency-insufficiency, then the relationship between

a differential word recognition score should produce differential values for comprehension performance.

Repeated performance on the same material produces less error on successive readings. (Powell, 1973; Busboom, 1973; Page, 1975; Kasdon, 1970; Lowell, 1970.) There is a reduction factor which the second and third reading of the material produces on the error effect. The second reading reduces error in word identification and comprehension about 25 to 33 percent; a third reading and other related practice further reduces error another 15 to 17 percent. Coupled with the error reducing effect of introducing vocabulary effectively prior to the first reading, it is apparent that effective teaching can transform a given piece of material from near frustration level to independent level. Therefore, a different set of evaluative criteria is necessary for diagnosis (oral at sight), a developmental lesson (silent before oral), and the completion process of teaching a given selection.

The Criterion Issue

The above characteristics of reading behavior should be reflected in any given set of criteria for evaluating reading performance. It is obvious that a uniform criterion cannot be applicable to the conditions of the word recognition and comprehension process. A differential set of values is indicated for both word recognition and comprehension scores.

The Betts (1946) criteria for the IRI is by far the most widely used values for evaluating reading behavior. While the Betts criteria is applicable for a given grade range (fourth) and under developmental con-

ditions at that level, it obviously cannot evaluate reading performance for the large number of readers outside that range or under different conditions. A differential set of criteria is essential for different levels and different conditions of reading. One set no longer will apply. Yet when one reads the descriptive accounts on the IRI in recent scholarly journals or newly released textbooks in reading, a traditional one non-differential set of criteria continues to be promulgated. As the frontier editor says in the film The Man Who Shot Liberty Valance, "When there's a conflict between the facts and the legend, we print the legend."

To effectively evaluate reading behavior informally, three pieces of information must be known and used for a reliable and valid measure: (1) what to include in effecting the criteria; (2) what not to include in the primary or first-order evaluation; and (3) what baseline criteria from which to form one's decision about the quality of reading.

In evaluating quality of oral reading, the evidence (Dunkeld, 1970) suggests that only errors of insertions, omissions, mispronunciations, substitutions, unknown words, and transpositions be included in determining the word recognition error ratio. Errors of repetition, punctuation omissions, self-corrections, proper names and places, dialect, and speech defects should be treated as symptomatic type errors and not used to effect the word recognition score. They may be recorded for secondary analysis if need be, but they should not be included in determining the initial quality of oral reading. No assumption should be made that although the six scorable type errors are given equal weight, they are in terms of quality. Obviously, they are not of equal importance, but as yet no

satisfactory method has been discovered to adequately weigh them properly.

Symptomatic type errors should not be used to effect the determination of word recognition accuracy. Fluency errors and observable behavior errors while distracting and probably desirable behaviors to apply extinction to, are not primary determiners in assisting in predicting the correct level of placement into instructional material.

The third essential item of information for measuring reading informally is to have some framework or guidelines by which to decide when a reading performance is moving toward the unsatisfactory end of the reading continuum. Criteria can be applied along three dimensions: comprehension, word recognition, and symptomatic behaviors. These dimensions are important, in that order, for determining the final decision as to the quality of reading. A discussion of each criterion dimension deserves attention.

The comprehension dimension is the most important of the three in determining final placement. When comprehension drops below the criterion for a given level, then it matters not what happens on the other two dimensions. The reading should end and a final judgement can be rendered. The comprehension questions asked following a reading (oral or silent) should be varied according to type. There should be questions on a literal level, implied types of questions, questions on the vocabulary used in that passage, and questions requiring evaluative skill. In all instances, the questions should be context dependent, i.e., answerable only by the information base supplied in the passage and dependent upon it. If the questions can be answered solely by prior experience or knowledge, then reading comprehension in that selection is not to have been measured.

The criteria for comprehension by clustering reading levels is given in Table 1. There is a differential effect by level and by condition. In materials at the readability levels of grades one and two, a 55 percent

Place Table 1 About Here

criterion is adequate for instructional placement when reading oral at sight (diagnostic function). For difficulty levels three, four, and five, 60 percent is the minimum satisfactory level, and for readability levels six or higher, a 65 percent criterion is essential to maintain quality of performance.

These data are derived by interacting cloze criterion scores from the Bormouth data (1972), the post oral cloze scores of Page (1975), the maximum cloze reconstruction figures (Weaver, 1977), with the word recognition figures given also in Table 1. (Powell, 1973.) While these scores are quite different from the traditional 75 percent criterion (which was derived from early school boards' criterion for examining and licensing teachers, i.e., arbitrarily determined), they are more realistic than the arbitrary percentages set without any particular precedent. A cloze score base has been determined by readers' actual performance. Extrapolating them into a differential percentage framework provides a more realistic evaluation of comprehension performance.

The word recognition criteria given in Table 1 has previously been published and discussed (Powell, 1970; 1971; 1973; and 1976). While the criteria is given in readability clusters (1-2, 3-5, 6+) for ease of

TABLE 1

Informal Reading Inventory
Scoring Criteria By Performance Level
and Condition

	Diagnosis		Developmental Teaching		Lesson Evaluation	
	W/R	Comp.	W/R	Comp.	W/R	Comp.
Independent Level						
1 - 2	1/17+	80+	1/17+	80+	1/17+	80+
3 - 5	1/27+	85+	1/27+	85+	1/27+	85+
6+	1/35+	90+	1/35+	90+	1/35+	90+
Instructional Level						
1 - 2	1/8 - 1/16	55-80	1/12 - 1/16	70-80	Converted to	
3 - 5	1/13 - 1/26	60-85	1/20 - 1/26	75-85	Independent	
6+	1/18 - 1/35	65-90	1/26 - 1/36	80-90	Level	
Frustration Level						
1 - 2	1/7 -	55-	1/11 -	70-		
3 - 5	1/12 -	60-	1/19 -	75-		
6+	1/17 -	65-	1/25 -	80-		

application in practice, it should be remembered that each level of difficulty has its own distinctive criterion figure (See Powell, 1970). The readability clusters are for convenience only of closely related figures. There is a distinct break in the pattern of scores between readability levels two and three and again between levels five and six. These natural breaks in reading growth form the cluster patterns given in Table 1.

When symptoms of difficulty are of value (in grades one and two), one or two symptoms are probably representative of the instructional zone for those children. More symptoms present than that are likely to represent the frustration level of reading while no observable symptoms would represent the independent reading level.

The initial task of measuring reading informally is to place the reader into material of the correct difficulty or readability level. This decision is first based on quantitative data. Then an error analysis (McLeod, 1918; Monroe, 1932; Goodman and Burke, 1972) is performed to make a qualitative analysis. Perfetti, Bell, and Goldman (1976) make the general conclusion "that the differences connected with reading skills are quantitative rather than qualitative." (p. 12) The issue is not one of either-or, but a matter of the purpose served, the need for additional information on reading behavior, and the order by which the data on a given reading performance is processed. Placement precedes analysis; quantitative precedes qualitative, but both are necessary for an effective diagnosis of reading beha

REFERENCES

Betts, Emmett A. Foundations of Reading Instruction. New York: American Book Co., 1946.

Bormuth, J. R. Literacy in the Classroom. In W. D. Page (ed.), Help for the Reading Teacher: New Directions in Research. Urbana, Ill.: National Conference on Research in English, 1975. pp. 60-90.

Busboom, J. B. An Investigation of the Relationship Among Various Testing Techniques of An Informal Reading Inventory at Instructional and Frustration Level. Unpublished doctoral dissertation, University of Illinois, 1974.

Carroll, J. B., Davies, P., and Richman, B. Word Frequency Book. Boston: Houghton-Mifflin Company, 1971.

Clay, M. M. Reading: The Patterning of Complex Behavior. Auckland, New Zealand: Heinemann Educational Books, 1972.

Dunkeld, C. G. The Validity of the Informal Reading Inventory for the Designation of Instructional Reading Levels: A Study of the Relationships Between Children's Gains in Reading Achievement and the Difficulty of Instructional Materials. Unpublished doctoral dissertation, University of Illinois, 1970.

Durrell, Donald D. Durrell Analysis of Reading Difficulty. New York: Harcourt, Brace and World, Inc., 1955.

Ehri, L. C. Effects of Word Recognition Training in a Picture-Word Interferance Task: Automaticity vs. Speed. Paper read at American Educational Research Association, Toronto, 1978.

Gilmore, John V. Gilmore Oral Reading Test. New York: Harcourt, Brace and World, Inc., 1968.

Goodman, Y. M. and Burke, C. L. Reading Miscue Inventory: Manual Procedure for Diagnosis and Evaluation. New York: The Macmillan Co., 1972.

Gray, William S. Gray Oral Reading Tests. H. M. Robinson, (ed.). Indianapolis: The Bobbs-Merrill Co., 1963.

Johns, J. L. Informal Reading Inventories: A Survey Among Professionals. Paper read at the International Reading Association, Anaheim, 1976.

Kasdon, L. Oral Versus Silent-Oral Diagnosis. In D. DeBoer (ed.), Reading Diagnosis and Evaluations. Newark, Del.: International Reading Association, 1970. pp. 86-92.

Killgallon, Patsy A. A Study of Relationships Among Certain Pupil Adjustments in Language Situations. Unpublished doctoral dissertation, The Pennsylvania State University, University Park, 1942.

Kinsbourne, M. Looking and Listening Strategies and Beginning Reading. In J. T. Guthrie (ed.), Aspects of Reading Acquisition. Baltimore: The John Hopkins University Press, 1976. pp. 141-161.

Lowell, R. Problems in Identifying Reading Levels with Informal Reading Inventories. In W. Durr (ed.), Reading Difficulties: Diagnosis, Correction, and Remediation. Newark, Del.: International Reading Association, 1970. pp. 120-125.

McLeod, L. S. The Influence of Increasing Difficulty of Reading Material Upon Rate, Errors, and Comprehension in Oral Reading. The Elementary School Journal, 1919, 17, 523-532.

Markman, E. M. Realizing That You Don't Understand: A Preliminary Investigation. Child Development, 1977, 48, 986-992.

Matejcek, Zdenek. Dyslexia in Czechoslovakian Children. In L. Tarnopol and M. Tarnopol (ed.), Reading Disabilities: An International Perspective. Baltimore: University Park Press, 1976. pp. 131-154.

Monroe, M. Children Who Cannot Read. Chicago: University of Chicago Press, 1932.

Pace, A. J. The Influence of World Knowledge on Children's Comprehension of Short Narrative Passages. Paper read at American Educational Research Association, Toronto, March, 1978.

Pace, A. J. and Golinkoff, R. M. Relationship Between Word Difficulty and Access of Single-Word Meaning by Skilled and Less Skilled Readers. Journal of Educational Psychology, 1976, 68, 760-767.

Page, W. D. The Post Oral Reading Cloze Test: New Link Between Oral Reading and Comprehension. Journal of Reading Behavior, 1975, 7, 383-389.

Perfetti, C. A., Bell, L., and Goldman S. Memory During Oral and Silent Reading. Paper read at the American Educational Research Association, 1976.

Powell, W. R. Reappraising the Criteria for Interpreting Informal Inventories. In D. DeBoer (ed.), Reading Evaluation. Newark, Del.: International Reading Association, 1970. pp. 100-109.

Powell, W. R. The Validity of the Instructional Reading Level. In R. E. Leibert (ed.), Diagnostic Viewpoints in Reading. Newark, Del.: International Reading Association, 1971. pp. 121-133.

Powell, W. R. Acquisition of a Reading Repertoire. Library Trends, 1973, 22, 177-196.

Powell, W. R. Monitoring Reading Behavior: Criteria for Performance. Research in Education, August, 1977, ERIC, ED 120 765.

Powell, W. R. and Dunkeld, C. A. Validity of the IRI Reading Levels. Elementary English, 1971, 2, 637-642.

Schummers, John L. Word Pronunciation in the Oral Sight-Reading of Third Grade Children. Unpublished doctoral dissertation, University of Minnesota, Minneapolis, 1956.

Spache, George D. Diagnostic Reading Scales. Monterey, California: California Test Bureau, 1963.

Steuver, R. Analysis of Relationship of Length of Passage to Categories of Oral Reading Errors. Unpublished doctoral dissertation, Olkahoma State University, 1967.

Watkins, M. A Comparison of the Reading Proficiencies of Normal-Progress and Reading Disability Cases of the Same I.Q. and Reading Level. Unpublished doctoral dissertation, University of Minneapolis, 1953.

Weaver, W. W. and Bickley, A. C. Some Differences in Encoding and Decoding Messages. In A. J. Kingston (ed.), Toward a Psychology of Reading and Language: Selected Papers of Wendell W. Weaver. Athens, Georgia: The University of Georgia Press, 1977. pp. 128-134.

Wohlwill, J. From Perception to Inference: A Dimension of Cognitive Development. In W. Kessen and C. Kuhlman (eds.), Thought in the Young Child. Monograph for the Society for Research in Child Development, 1962, No. 83, 27:87-112.