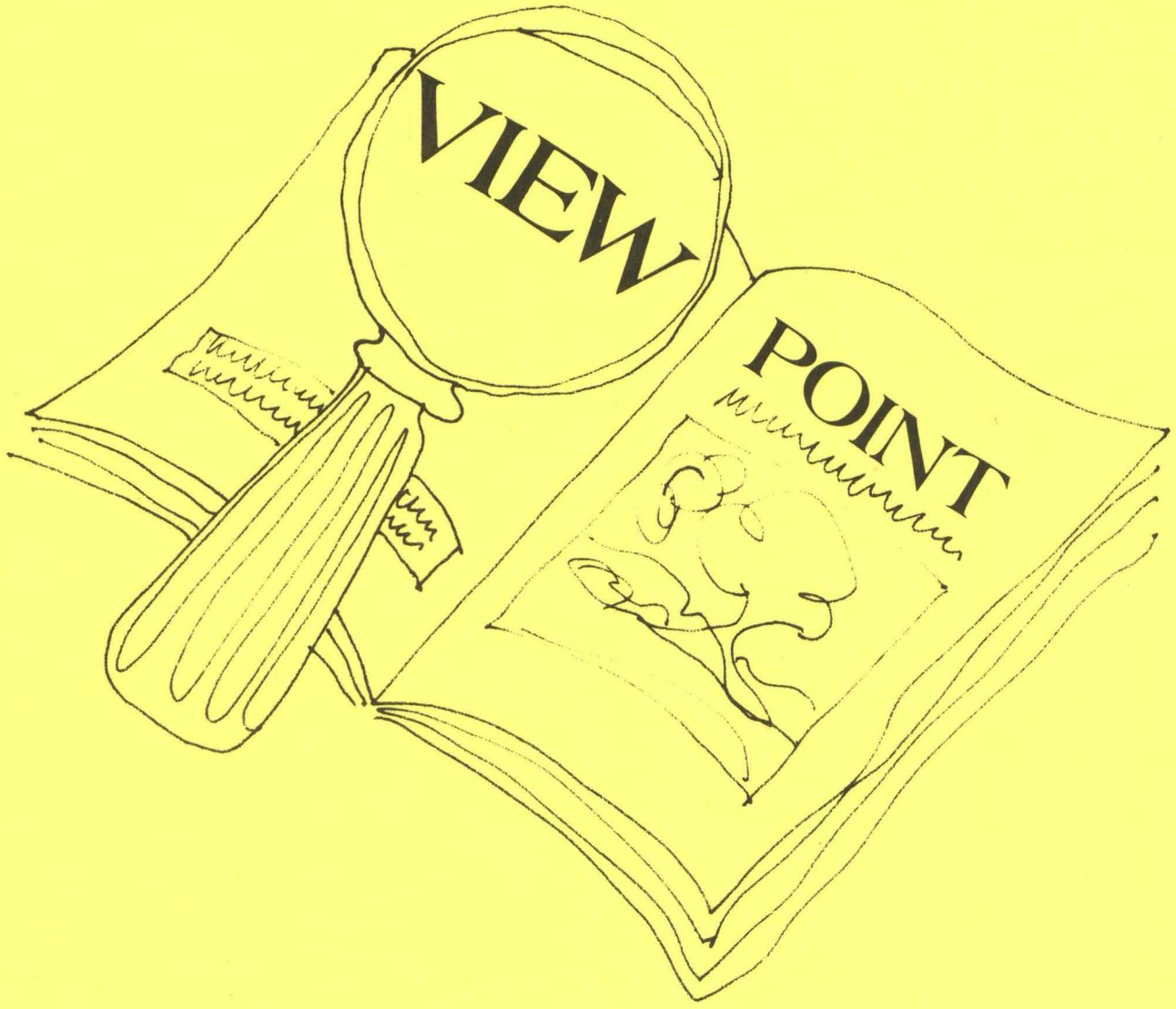
Analysis of Reading Problems: A School Psychologist's



Iowa Department of Public Instruction

Analysis of Reading Problems: A School Psychologist's View Point

By Portia Blackman

State of Iowa
DEPARTMENT OF PUBLIC INSTRUCTION
Grimes State Office Building
Des Moines, Iowa 50319

STATE BOARD OF PUBLIC INSTRUCTION

Jolly Ann Davidson, President, Clarinda
John E. van der Linden, Vice-President, Sibley
Cornelius Bodine, Jr., Sioux City
William K. Cropp, Des Moines
Karen K. Goodenow, Wall Lake
Virginia Harper, Fort Madison
Robert G. Koons, Clinton
Susan Wilson, Waterloo
Harold R. Yeoman, Monticello

ADMINISTRATION

Robert D. Benton, State Superintendent and Executive Officer
of the State Board of Public Instruction
David H. Bechtel, Administrative Assistant
James E. Mitchell, Deputy Superintendent

Pupil Personnel Services Branch

Drexel D. Lange, Associate Superintendent
J. Frank Vance, Director of Special Education
Jeff Grimes, Consultant Psychological Services

INTRODUCTION

To the school psychologist reading is more than a score on a test. It is a complex process. In the text which follows Portia Blackman, the author, makes it clear that school psychologists have an array of approaches and assessment alternatives available in order to analyze reading performance. A complete, competent and meaningful analysis of reading requires time, careful consideration of reading errors and a review of the educational materials the student will be using in the classroom. Portia's suggestions regarding how to analyze dysfunctional reading, determine the student's instructional level and make the most of out of assessment data are refreshing and should prove useful to all school psychologists.

Portia Blackman is a school psychologist in AEA 13 (Halverson Center for Education. R.R. #1, Council Bluffs, Iowa 51502). She is beginning her fourth year as a psychologist in Iowa, and has practical experience in the analysis and diagnosis of reading problems from a school psychologist's point of view. Her master's degree is from the University of Iowa from the Institute of Child Behavior and Development and specialist degree is in School Psychology also from the University of Iowa.

I think school psychologists will find this text to be informative and useful in providing a quality analysis of reading problems presented by children with special needs.

Jeff Grimes
Consultant for School Psychological Services
Iowa Department of Public Instruction
Grimes State Office Building
Des Moines, Iowa 50319

ANALYSIS OF READING PROBLEMS: A SCHOOL PSYCHOLOGIST'S VIEWPOINT

by Portia Blackman

School psychologists often tend to deal with the diagnosis of reading problems in a superficial manner. The following scenario depicts one all-too-common approach to children's reading difficulties on the part of school psychologists. The psychologist administers a battery of tests to a student, including an individual intelligence test, a reading test, typically the Wide Range Achievement Test, and a test of visual-perceptual functioning, such as the Bender Visual-Motor Gestalt Test. In an effort to be more thorough, the psychologist also may administer an individual diagnostic reading test, such as the Gilmore Oral Reading Test or the Spache Diagnostic Reading Scales. The psychologist then uses the obtained IQ score to determine the student's expected achievement level, and he uses the grade equivalent score obtained on the reading test as an indication of the student's current level of reading achievement. The comparison of expected and observed achievement levels is taken as an indication of the extent of the student's underachievement in reading and often is used in making program placement decisions (e.g., placement in a learning disabilities resource program). If a student is found to be underachieving in reading and also does poorly on the test of visual-perceptual skills, the psychologist often infers a causal relationship between the student's poor visual-perceptual functioning and his poor reading achievement.

The grade equivalent score obtained on the reading test is assumed to represent the grade level of reading materials in which the student should be placed for reading instruction. Thus, the psychologist may write in his report, "Jimmy's intellectual ability indicates that his

expected reading achievement is at the 5.4 grade level, whereas his actual reading achievement, as measured by the WRAT, is at the 3.4 grade level." On this basis, he may recommend participation in a learning disabilities resource program due to significant underachievement in reading, and he either states or implies that Jimmy's reading instruction should be in materials at the 3.4 grade level. He then may go on to state that Jimmy's reading problems are due to his deficit in visual-perceptual skills, as evidenced by his poor performance on the Bender. With this, his job is finished, and he moves on to the next referral.

The approach described above, though very common and often accepted as sufficient, is actually inaccurate and inadequate for several reasons, some of which can be stated briefly at the outset. (1) The student's IQ score may not be the most accurate basis on which to determine his reading potential or his expected reading achievement. (2) The obtained grade equivalent reading score may not represent the student's actual level of reading achievement at all, because of the manner in which the reading test assesses reading and because the scores obtained are usually based on normative data. (3) Even if the obtained grade equivalent reading score were accurate, it could not be used directly to select the appropriate instructional materials for the student. (4) The score the student obtains on the reading test provides no clues as to what instructional procedures would be most beneficial for the student. (5) The relationship between poor performance on visual-perceptual tests and low reading test scores may be more coincidental than causal. Each of these shortcomings will be discussed in detail in this paper, and various solutions will be offered.

Whenever a child's reading ability is evaluated, the school psychologist should strive to do more than determine that student's score on a standardized reading test. The important goal to strive for is to

provide a student with the specific instructional materials and procedures he needs in order to make maximum progress in reading. Many school psychologists may feel that this is someone else's job -- maybe the classroom teacher's, the resource teacher's, or the consultant's. In many schools, it is true, one or more of these professionals may take this responsibility upon themselves, but there are other schools where no one is performing this crucial function. Whether due to lack of expertise or lack of concern, the job is not getting done.

School psychologists should not assume that someone else will translate their test scores into optimal instructional materials and procedures for each student. We need to take responsibility for this ourselves, when necessary, and we cannot do this unless we familiarize ourselves with the tools for turning test scores into the most appropriate instructional techniques for each student we evaluate.

This paper will present and discuss a variety of tools and techniques that school psychologists can employ to accurately assess each student's reading skills and to specifically determine what materials and methods should be used to promote maximum progress in learning to read or overcoming reading difficulties.

Estimating Reading Expectancy

One question of some importance in evaluating a student with reading difficulties is at what level he should be able to read. The usual way this judgment is made is to consider that the student should have reached a reading age or grade level comparable to his mental age or grade level. But the assumption that a student should be achieving up to his mental age is a shaky one, because there probably is not a perfect correspondence between intelligence and reading ability. Spache

(1978, p. 131) cites several studies indicating that the statistical relation between reading and mental age is a very moderate one at primary levels and only later increases with grade level to a substantial level.

There are many variables besides mental ability that influence a child's potential reading ability. The child's motivation, the amount and quality of his schooling, his familiarity with standard English, the adequacy of his socioeconomic and cultural background, his verbal ability, and his abilities in nonreading areas such as mathematics all contribute in some way to the level of reading ability he could be exptected to attain. Nevertheless, the child's mental age, as determined by an intelligence test, is generally thought to be the best single predictor of reading expectancy.

Some formulas use mental age alone as the basis for estimating a child's reading expectancy. Kolson and Kaluger (1963) determine a Learning Expectancy Level (L.E.L.) which they say indicates the grade level at which a child may be expected to learn to read, all other factors being normal. Their formula is L.E.L. equals mental age minus 5. Harris (1971) mentions this procedure and recommends subtracting 5.2 instead--the 5 or 5.2 representing the number of years prior to school entrance. However, Harris no longer feels that such a formula is adequate because it ignores many factors affecting reading ability. He has proposed the use of an Expectancy Age in which mental age is given twice the weight of chronological age (Harris, 1970):

$$\text{Expectancy Age} = \frac{2\ MA + CA}{3}$$

His rationale as he stated it is: "Chronological age is the dimension in which growth and learning take place; and when it is used in computing an expectancy score, it can represent a composite of the many factors besides intelligence that can influence a child's growth

in reading competence" (Harris, 1971). Spache (1978) points out that Harris' formula is precisely the same as that offered by Alfred S. Lewerenz in 1939 (Lewerenz, 1939).

Bond and Tinker (1967) also object to the use of mental age as the sole criterion for determining reading expectancy. They find fault with what they call the Mental Grade formula, which they present as: Mental Grade = (CA x IQ - 6.2) +1.0, where 6.2 is the typical age at which a child enters first grade. It can be seen that this formula is no different than that mentioned by Harris (1971): Mental Grade = MA - 5.2. The problem with the Mental Grade formula, according to Bond and Tinker, is that it overestimates reading expectancies for children with IQs two or more standard deviations above average, and it underestimates reading expectancies for children with IQs two or more standard deviations below the norm. That is, too much is expected of bright children in terms of reading ability and too little is expected of slower children. For instance, a child with an IQ of 170 could not realistically be expected to read at the eighth grade level after the first one and one-half years of school. Similarly, a child with an IQ of 60, they believe, would commonly be more advanced than nursery school by his second grade year.

Bond and Tinker (1967) proposed what we will call a Years in School formula:

Reading Grade = (IQ x Years in School) + 1.0

The 1.0 is added "...because the child starts school at grade 1.0 and after a year in school, the average child is at grade 2.0 or just entering second grade." The Years in School formula is based on the assumption that the IQ is an index of rate of learning new experiences.

The authors state, "On the whole, it can be assumed that the usual child with 150 IQ can be expected to learn new things, when presented, about one and a half times as fast as the average child. Likewise, the child with a 75 IQ can be expected to learn them only about three-fourths as fast as the average child." By this formula, the typical child with an IQ of 70 would be expected to read at the 1.7 grade level at the end of first grade and at 2.4 at the end of second grade. The child with 150 IQ could be expected to read at the 4.0 level at the end of second grade and at the 7.0 level at the end of fourth.

Bond and Tinker (1967) provide data indicating that their Years in School formula provides a much better fit with observed reading ability than does the Mental Grade formula, based on a study of 379 fifth-graders. They state, "...the formula applied at the fifth-grade level gives estimates that are startlingly close to the observed reading averages for almost every level of IQ."

Harris (1971) objects to Bond and Tinker's formula because he feels it provides expectancies that are much too high for mentally slow children. Young (1976) also finds shortcomings in the Years in School formula. Although she finds it much more accurate than the Mental Grade formula at the extremes of intellectual ability, she notes that it functions unsatisfactorily for children who have been retained one or more years (as does the Mental Grade formula). Children who repeat a grade do not make a normal year's growth in achievement during that year, she notes, and she cites research supporting this assertion. Young also mentions the ease with which errors can be made using Bond and Tinker's formula.

Young (1976) offers a formula which "...avoids distortions and

reduces the chances for computational errors." It is calculated by multiplying the child's present grade by his IQ score, written as a decimal. The formula is:

Grade x IQ = Reading Expectancy Level

Thus, if a child is in the middle of the fourth grade with an IQ of 80, his Reading Expectancy Level is 4.5 x .80 = 3.6. A table presented by Young (1976) is reproduced here to illustrate the comparison among the Mental Grade, Years in School, and Young formulas. (see Table 1).

Table 1. Comparison of three formulas for estimating reading expectancy*

	Second grade (February mean age 8.0)					
	Young Formula		Years in School Formula		Mental Grade Formula	
IQ	Social promotion	Retained one year	Social promotion	Retained one year	Social promotion	Retained one year
170	4.3	4.3	3.6	5.3	8.1	9.8
150	3.8	3.8	3.3	4.8	6.5	8.0
130	3.3	3.3	3.0	4.3	4.9	6.2
110	2.8	2.8	2.7	3.8	3.3	4.4
100	2.5	2.5	2.5	3.5	2.5	3.5
80	2.0	2.0	2.2	3.0	K.9	1.7
60	1.5	1.5	1.9	2.5	N.3	K.1

	Sixth grade (February mean age 12.0)					
	Young Formula		Years in School Formula		Mental Grade Formula	
IQ	Social promotion	Retained one year	Social promotion	Retained one year	Social promotion	Retained one year
170	11.1	11.1	10.4	12.1	14.9	18.3
150	9.8	9.8	9.3	10.8	12.5	15.5
130	8.5	8.5	8.2	9.5	10.1	11.4
110	7.2	7.2	7.1	8.2	7.7	8.8
100	6.5	6.5	6.5	7.5	6.5	7.5
80	5.2	5.2	5.4	6.2	4.1	4.9
60	3.9	3.9	4.3	4.9	1.7	2.3

* Reproduced from Young, 1976

As can be seen in Table 1, the reading level expectancies obtained with the Young formula and the Years in School formula are not so extreme for students with very high or very low IQs as are those obtained with the Mental Grade formula. The main difference between the Young and Years in School formulas has to do with reading expectancies for students who have been retained one or more years in school.

This psychologist cannot recommend which formula to use among those presented, and it may be that better predictions can be made with a formula that is not mentioned above. However, it seems important to use one of these formulas when evaluating any student, in order to determine whether a student's low reading achievement score reflects an actual reading disability or a general lack of intellectual ability. It does seem advisable, however, to avoid the mental age or mental grade formula when predicting reading potential for brighter children. This formula is too likely to label a bright child as reading disabled even though his reading ability may be above his grade level.

Searching for Causes of Reading Problems

When a student's reading achievement is found to be significantly below his reading potential or his reading expectancy level, school psychologists are fond of trying to identify the "cause" of the student's reading problem. The real causes of reading problems are probably incredibly varied and complex, and a school psychologist cannot be expected to determine the causes of a child's reading difficulties in a single session with him. Nevertheless, it is tempting to search the child's history, his behavior, and his test performance for some clue as to the cause of his poor reading skills.

The relationship of certain visual-perceptual abilities and school

learning has been of great interest to educators, many of whom have speculated that deficits in visual perception may cause or, at least, contribute to academic problems. Historically, this assumption of a causal relationship has been based on the clinical experience of various educators as well as on some basic research. One result of this assumption is that some school psychologists have classified students as "visually perceptually handicapped" on the basis of their performance on geometric figure copying tests such as the Bender Visual-Motor Gestalt Test or the Beery Visual-Motor Integration Test and have then attributed their failure to learn to read adequately to their problems in visual perception.

Larsen and Hammill (1975) point out that the educational usefulness of the relation of visual perception to school learning has never been fully substantiated. In fact, they state, recent opinion and research have seriously questioned the presumed causal relation between visual perceptual problems and school learning problems (Cohen, 1969; Hammill, 1972; Mann, 1970). Meanwhile, many school systems are committing large numbers of hours and dollars to "train" visual-perceptual skills as a means of ameliorating learning problems. Larsen and Hammil state that before visual-perceptual training can be considered justified, its relationship to academic achievement must be established.

Larsen and Hammill (1975) reviewed the research exploring the relationship of visual discrimination, spatial relations, visual memory, and auditory-visual integration to school learning. They reviewed a total of 60 studies that employed a variety of tests of visual-perceptual skills, including the five subtests of Frostig's Developmental Test of Visual Perception and the Bender Visual-Motor Gestalt Test, among several

others. The reader is referred to Larsen and Hammill (1975) for the details of the analyses they performed on the findings of these studies; only the conclusions will be presented in this paper. The authors state, "The combined results of the correlational research treated in this paper suggest that measured visual-perceptual skills are not sufficiently related to academic achievement to be particularly useful." They found that, when the variable of intelligence is accounted for, "...children who do poorly in reading do not differ in visual-perceptual ability from children who read at age expectancy."

Larsen and Hammill go on to say, "Apparently, a large percentage of children who do adequately on tests of visual perception experience difficulty in school learning, and an equally sizable percentage who do poorly on these same tests exhibit no problems in school learning. As the relationship of visual perception to school learning is, at best, minimal, the time and expense currently devoted to visual training in the schools should be reevaluated if the purpose of such training is to improve academic achievement."

A more recent study by Fisher and Frankfurter (1977) tested three hypotheses about visual-perceptual difficulties and found that none of them seemed to adequately explain reading disabilities. They concluded, "These data discredit perceptual difficulties as the cause of reading disability are a developmental lag which is cognitively based, at the level of naming or translating graphological to phonological information, rather than perceptually based." This hypothesis needs a great deal of research before it can be widely accepted and utilized.

In the meantime, on the basis of recent research, it seems advisable that school psychologists abandon their efforts to attribute reading

difficulties to deficits in visual perception. Rather than trying to establish the "cause" of a student's reading problems, it would be more educationally useful to concentrate on the specific reading skills the student does and does not demonstrate. A detailed and insightful analysis of how the student reads would be more helpful to those working with him than would a statement as to the cause of his deficiencies in reading.

The Use of Standardized Diagnostic Reading Tests

In evaluating a student's academic achievement, some school psychologists are satisfied with administering only one individual achievement test, such as the Wide Range Achievement Test (WRAT) or the Peabody Individual Achievement Test (PIAT). While this practice is certainly efficient, it does not yield much information about the student's academic performance, particularly in reading. Neither the PIAT nor the WRAT provide the kind of rich diagnostic information that can be obtained simply by having a student read a few sentences aloud.

Many school psychologists have taken a step in the right direction by including a diagnostic reading test among their evaluation techniques. Among the tests they may use are the Gray Oral Reading Tests, the Gilmore Oral Reading Test, the Gates-McKillop Reading Diagnostic Tests, the Durrell Analysis of Reading Difficulty, and Spache's Diagnostic Reading Scales. These tests usually include a series of graded paragraphs for oral and silent reading, followed by comprehension questions. The student's oral reading of these paragraphs can reveal much valuable information about his reading abilities. Some of the tests provide graded lists of words to use in determining the extent and level of the student's sight vocabulary and his skills in word analysis. Some

of these tests also may include special sections designed to reveal information about reversals, word blending, and other specific word analysis skills. Thus, the wealth of information that such tests yield about how a student reads make them a highly valuable addition to the school psychologist's repertoire of testing procedures.

Limitations of Grade Equivalent Scores

Although the school psychologist may observe much useful information about a student's reading skills when he administers a diagnostic reading test, the report he writes may only refer to the grade-equivalent score the student earned on the test. Having reported the earned grade-equivalent score, the psychologist may go on to state that the student should be placed in instructional materials at that grade level. If this is not stated, it may be assumed by those reading the report, because it is a common misconception that the grade-equivalent score a child earns on a reading test represents the appropriate readability level for his instructional materials.

Actually, all the diagnostic reading tests mentioned above have derived their grade-equivalent scores from normative data. This means that a child who obtains a grade-equivalent of 2.5 on one of these tests is capable of reading as well as the average second-grader who was a member of the standardization group for that particular test. In most cases, however, the standardization group is not adequately described, so it is impossible to determine the readability level of instructional materials that would be appropriate for the average second-grader in the group. Consequently, it is also impossible to translate the child's earned grade-equivalent of 2.5 into a suggested readability level for his placement in instructional materials.

The following description of the scoring systems of several diagnostic reading tests should illustrate the problems inherent in grade-equivalents based on normative information. This information was obtained largely from an excellent paper by Joyce Hood at the University of Iowa. This undated manuscript, which shall be referred to hereafter as Hood (unpublished), deals eloquently with a number of the issues discussed in this paper.

Gray Oral Reading Tests. The grade-equivalent earned by a child who takes this test is based on the number of oral reading errors and the speed of oral reading. Comprehension is not reflected in the grade-equivalent score. The same grade-equivalent may be earned by children with qualitatively different oral reading abilities. For example, one child may be a relatively slow but accurate reader while another is a fast but inaccurate reader.

The grade-equivalents were derived from a standardization group of children with a mean intelligence quotient of 110. Thus, the grade-equivalent a child earns on this test, e.g. 3.1, means that the combination of his oral reading errors and speed resulted in a score which is the same as that earned by the typical child with high-average intelligence who is just beginning his third-grade year. In order to use the grade-equivalent as a guide to placement in instructional materials one must determine what readability level is appropriate for instructing the third-grader who has high-average intelligence.

Gilmore Oral Reading Test. This test yields separate scores for oral reading accuracy and comprehension and a rating of fast, average, or slow for the rate of reading. A child's grade-equivalent score in each area is based on the performances of the children who made up the

standardization population. Thus, a grade-equivalent of 7.8 in oral reading accuracy means that the child has made the same number of oral reading errors as the average child among the 280 seventh-graders who took that form of the test during the standardization administration. The school systems in the standardization program are said to represent a variety of socioeconomic backgrounds, but there is no evidence that they were proportionately representative of all socioeconomic levels. Neither is there a report of the average intelligence of the standardization population. This makes it difficult to estimate accurately what readability level is appropriate for the average seventh-grader in the standardization population. Thus, a student's grade-equivalent of 7.8 cannot be translated directly into an appropriate placement in instructional materials.

Gates-McKillop Reading Diagnostic Tests. The grade-equivalent earned by a child who takes this test is based on his oral reading errors only. It is derived from the performances of some unspecified standardization group of unknown size and origin. "Thus the pupil who gets a raw score of 15 does about as well as the average child at midyear in grade three" (Manual of Directions, page 2). If the authors had reported what type of community that average child resides in, and whether he possesses average learning ability, it might be possible to determine what readability level is appropriate for that average child to read, and then to try that level as an appropriate instructional placement for the child who raw score on this test is 15. Since the norming population is not described, the grade-equivalent cannot be used in this manner with any degree of confidence.

Durrell Analysis of Reading Difficulty. The child's rate of oral reading provides the only basis for the grade-equivalent he earns on this test. According to the Manual of Directions (page 32) the norms are

based on at least a thousand children who took the tests, but the socio-economic levels and intellectual abilities of these children are not described. Thus the grade-equivalent earned on this test corresponds to the average achievement of a large group of children of unspecified socioeconomic background and general learning ability. Since it is impossible to estimate the readability levels of instructional materials that would be appropriate for these children, the grade-equivalent a child earns on this test cannot be translated into an appropriate placement in instructional material.

Spache's Diagnostic Reading Scales. The instructional level on this test is the highest level at which the child's oral reading errors do not exceed the recommended maximum standard and his comprehension score does not fall below the minimum standard. These standards represent the average performance of pupils reading at this level according to the examiner's manual (page 28). As in all the previous tests except the Gray Oral Reading Tests, however, neither the socioeconomic levels nor the intellectual abilities of the children on whom these standards were based are described. In order to use the grade-equivalent a child has earned on this test as a guide for instructional placement, it would be necessary to know what readability level of materials has been found appropriate for the average child of that grade level in the standardization group.

The Use of Criterion-Referenced Oral Reading Tests

If grade-equivalent scores based on norms do not provide educationally useful information, perhaps these scores should not be utilized when the purpose of testing is to determine a student's level of reading proficiency and, therefore, the level of instructional materials appropriate for him.

For this purpose another type of test is needed--one that does not derive its scores from normative information.

Criterion-referenced oral reading tests base their scoring systems on certain predetermined criteria of performance, such as 95 percent correct word recognition and/or 75 percent correct comprehension. The most widely accepted standards for the criteria used by criterion-referenced reading tests will be discussed fully in a later section. The point here is that a student's score depends upon his ability to meet certain criteria for acceptable performance, rather than upon how his reading ability compares with that of a poorly described standardization group. Three criterion-referenced reading tests are discussed briefly below.

Standard Reading Inventory. The content of this test was designed to be representative of three basal reading series whose copyright dates range from 1948 to 1961. The manual states "...no words are used in the stories for primer through 3-2 levels which have not been introduced in two of the three basal reading series at or before the level of the story (page 41). The criteria McCracken uses for percentages of oral reading accuracy and comprehension correspond to criteria which he has recommended elsewhere (McCracken, 1967), and they are not based on normative data. Due to this and to the fact that McCracken suggests that an experienced examiner may choose to use his own standards, this test appears to be a criterion-referenced test.

This test includes graded vocabulary lists, graded paragraphs for oral and silent reading, and comprehension questions following each paragraph. Farr (1969) discusses the results of McCracken's efforts to establish the content validity, construct validity, and reliability of this test, and he concludes, "...it certainly appears that the Standard

Reading Inventory should validly determine students' functional reading levels."

Classroom Reading Inventory. This test uses scoring criteria that are similar to those recommended by McCracken (1967). Specifically how the oral paragraphs were prepared is not explained, although it is claimed that the paragraphs are similar to the type of reading material found in various grade levels throughout elementary school (Silvaroli, 1969, page xi). The readability levels of the paragraphs were evaluated using readability formulas. Because the criteria for determining reading levels are not based on normative data, this test is considered to be a criterion-referenced test. According to Hood (unpublished), the content validity of this test is not well documented.

Botel Reading Inventory. This inventory consists of tests of silent and oral word recognition abilities rather than tests of paragraph reading and comprehension. The words in the test lists through the high-third reader level were selected from the Botel study of 1185 Common Words (Botel, 1968), previously known as the Bucks County 1185 Common Words. The scoring criteria were developed in relationship at reader level placements of children who were performing at various reading levels according to criteria similar to those recommended by McCracken (1967). The test content and scoring criteria indicate that it is a criterion-referenced test.

It should be pointed out that the paragraphs in the Gray Oral Reading Tests and the Gilmore Oral Reading Test were prepared with the aid of lists of words and the grade levels at which these words first appeared in a sample of widely used readers. Although the grade equivalent scores on these tests are based on normative information, it would be possible

to score each test passage according to established criteria for acceptable performance. Thus, these tests could be used as criterion-referenced rather than norm-referenced tests.

Content Validity of Criterion-Referenced Tests

A criterion-referenced test is no better than the sample of test items included and how representative that sample is of the area to be tested. In order to determine the content validity of a criterion-referenced tests, one should demonstrate the comparability of its content to the content of instructional materials. This procedure requires a common standard of measurement, which suggests that a single readability measure should be applied to both the test and the instructional materials that are to be related to it.

Formulas for Predicting Readability

A readability measure is used to determine how readable a piece of writing is, in terms of how easily it can be comprehended by the reader. The readability of a book must be matched to the reading ability of the child who is to be instructed from that book if optimal learning is to take place. There are many ways of measuring the readability of a sample of writing. Judgments by readers (particularly professionals such as librarians) and comprehension tests are two fairly accurate ways of determining the readability of written material, but they are both too costly and time-consuming to be practical for most educators (Klare, 1974-75). Predicting readability, on the other hand, can be accomplished with relatively little effort by using a readability formula. This method uses counts of language variables in a piece of writing in order to provide an index of probable difficulty for readers (usually expressed in terms of grade levels). Following are brief descriptions of a few of

the most widely-used formulas for determining the readability of written materials.

Flesch's Reading Ease formula. Of the several formulas that Rudolf Flesch devised, the most well-known is his Reading Ease formula. He used the McCall-Crabbs Standard Test Lessons in Reading (McCall and Crabbs, 1925) as a criterion of difficulty, in that the predictive validity of his formula was based on its ability to predict the comprehension levels of children reading the McCall-Crabbs passages. In other words, the criterion to which his readability ratings were related was the grade placement at which a specified level of comprehension resulted on the McCall-Crabbs Lessons.

Flesch's Reading Ease formula (Flesch, 1948), which correlated .70 with the McCall-Crabbs criterion, is as follows:

R. E. (Reading Ease) = 206.835 - .846 wl - 1.015 sl

Where: wl = number of syllables per 100 words;
sl - average number of words per sentence

The Dale-Chall formula. Dale and Chall presented their formula for adult materials in 1948, and it quickly became, along with the Reading Ease formula, one of the two most widely used. It also used the 1925 McCall-Crabbs Standard Test Lessons in Reading as a criterion. The Dale-Chall formula, which correlated .70 with McCall-Crabbs criterion scores, is as follows:

$$X_{c50} = .1579x_1 + .0496x_2 + 3.6365$$

Where: X_{c50} = reading grade score of a pupil who could answer one-half the test questions on a passage correctly;
X_1 = Dale score, or percentage of words outside the Dale list of 3,000;
X_2 - average sentence length in words.

Powers, Sumner, and Kearl (1958) provided a recalculated version of the Dale-Chall formula, based upon the 1950 edition of the McCall-Crabbs

Lessons. They found a correlation with 1950 McCall-Crabbs scores of .71, which is virtually the same as the .70 found with the 1925 McCall-Crabbs scores by Dale and Chall (1948). This, plus other consistent evidence, suggested that the Dale-Chall was the most accurate general-purpose formula available up to 1960 (Klare, 1963). The recalculated formula is:

$$X_{c50} = 3.2672 + .0596\ X_2 + .1155\ x_1$$

After a new version of the McCall-Crabbs Lessons appeared in 1961, Holmquist (1969) recalculated the Dale-Chall formula as follows:

$$X_{c50} = .0512\ X_2 + .1142\ x_1 + 3.442$$

This formula had a correlation of .69 with the criterion.

The Spache formula. George Spache developed his original formula for children's material of grades 1 to 3 in 1953.

$$\text{Grade level} = .141x_1 + .086x_2 + .839.$$

Where: x_1 - average sentence length in words;
X_2 - number of words outside the Dale list of 769 words.

Spache validated his formula against level of classroom use for 152 books in grades 1 to 3, finding a multiple correlation coefficient of .818.

Hood (unpublished) states that her experience with the Spache formula suggests that it does not discriminate effectively below the third-grade difficulty level. Since it uses an ungraded list of 769 words ("the Dale Easy Word List"), the primary factor operating at earlier grade levels is the factor of average sentence length.

Stone (1956) found that by using his revision of the word list, Spache's formula yielded lower readability ratings than with the original Dale list. Spache subsequently followed this procedure in using Stone's Revised Word List (Spache, 1966). Recently, Spache (1974) revised his formula using the Harris-Jacobson Basic Elementary Reading Vocabularies.

He has reported a correlation of .95 between formula scores and grade level of primary books.

The Botel formula. The Botel method of predicting readability is not a regression equation as most others are. Botel's method (Botel, 1962) is to predict readability level from the median difficulty of samples of words whose grade levels are determined through their presence in or absence from a "Graded Vocabulary List" (based on the "1185 Common Words" in Botel, 1968). The obtained Botel Readability Score may vary from preprimer to grade 12 in difficulty. Botel validated his formula by comparing the vocabulary with that used at various levels of basic readers; in books used in grades 5 to 60; in junior and senior high school textbooks; and in Reader's Digest, Time, and the New York Times. The usefulness of this formula will be discussed in more detail in later sections of this paper.

Fry's Readability Graph. Fry (1968) proposed a "Readability Graph" for predicting readability, recommending it as a way of saving time and effort. Fry used the factors of syllables per 100 words and words per sentence. The user simply enters the counts of these two factors in a graph and reads the readability score directly from the graph. Fry's graph has been validated on both elementary and secondary materials, and the scores derived from it correlate highly with those from several well-known formulas. It correlates .94 with the Dale-Chall formula and .96 with the Flesch formula, though only .76 with the Botel formula, which estimates higher readability levels. Maginnis (1969) has recently extended the Fry graph downward through the preprimer level.

Choosing a Readability Measure

There are many other formulas available for predicting readability

than the few mentioned above. The potential user is faced with so many formulas to choose from that the choice may be very difficult. Klare (1974-75), in his excellent review of formulas and revisions since 1960, suggests some general guidelines for choosing a readability formula. He states, "...there is little to be gained from choosing a highly complex formula. A simple two-variable formula should be sufficient, especially if one of the variables is a word or semantic variable and the other is a sentence or syntatic variable."

Klare (1974-75) goes on to state, "The word or semantic variable is consistely more highly predictive than the sentence or syntactic variable when each is considered singly. When a word variable is to be counted, there are two common choices: count word length, or count number of words not on a particular list of familiar words. Using a list of familiar words appears to give a slightly more predictive index than counting word length, probably because length is a secondary reflection of familiarity.

"The sentence variable, though not as predictive of difficulty as the word variable, does have an important contribution to make to formulas. Though sentences can be evaluated in several ways, a simple count of length is generally sufficient. Sentence compexity is probably the real casual factor in difficulty, but length correlates very highly with complexity and is much easier to count."

The Botel method of predicting readability differs from the others in the fact that it is based on a word or semantic variable only (that of word familiarity), and it utilizes a graded vocabulary list. Hood (unpublished) has found that the Botel method predicts higher readability levels than do many other formulas, but she feels it is a superior predictor at the primary grade levels. Since the Botel approach is based on a word recognition criterion rather than a comprehension criterion,

it is well-suited to beginning readers, for whom word recognition is the major difficulty.

However, Hood has discovered problems with the Botel method, which she describes as a "word recognition level readability estimate" (as opposed to a comprehension level readability estimate, as are most others). Due to the increasing vocabulary load in the basal readers of the 1960's and 1970's, the graded vocabulary lists used in Botel's Predicting Readability Levels (1962) might not include some words that are now considered highly familiar at particular grade levels; thus, the readability estimates the Botel method provides might now be too high. This drawback also may apply to other formulas that utilize lists of familiar words in predicting readability.

The word recognition level readability estimate also does not reflect the contribution of context and picture clues in word recognition, as Hood (unpublished) points out. For instance, the word "mew" is considered of fourth-grade difficulty in the Botel lists; however, the word "mew" may be easily recognized by younger children if it appears in the context of a story about kittens, particularly if the text includes a picture of a kitten. Hood (unpublished) states, "It is assumed, but it has not been proved, that the presence of these additional cue sources compensates for the increased vocabulary load and that the newer readers are not any more difficult to read." If so, the Botel method may still be as accurate as it was for the older basal readers on which it was based.

However, Hood (unpublished) feels that readability estimates such as the Botel are inappropriate to the newer readers in which interesting pictures and memorable phrases may contribute to reading ease, as in the Sounds of Language readers by Holt, Rinehart, and Winston. Hood also feels that readability estimates based on lists of familiar words are

not appropriate for use with materials based on a controlled spelling pattern approach to reading instruction.

Variations in Publisher's Estimates of Readability

Now that readability measures have been discussed, one of the main points of this paper can be brought up again: that is the importance of applying the same readability formula to both the reading test you administer and the instructional materials the student will be using, in order to arrive at the best level of instructional materials for that student. One reason that the same formula should be applied to both test materials and instructional materials is that there is such wide variability among different tests and books in terms of their readabilities.

Spache (1978) refers to a study he did several years ago of the readability levels of all the basal readers in the twenty basal series then on the market. Although he does not say what readability formula(s) he used, he mentions that each book was analyzed for readability level at five to ten points scattered throughout the book and those samples were averaged for each book. Table 2 presents the results of this comparative study. Spache (1978) states, "It is apparent that basals offered by different publishers for the same grade level vary greatly in their actual reading level.

Table 2. Readability Levels of Basal Readers*

Level	Number of Books	Range	Median
Preprimer	44	1.1 to 2.2	1.5
Primer	18	1.2 to 2.9	1.8
First	22	1.8 to 3.1	2.0
Second	30	2.0 to 4.2	2.6
Third	32	2.5 to 4.8	3.4
Fourth	17	3.5 to 5.1	4.4
Fifth	16	4.5 to 5.9	4.9
Sixth	16	4.5 to 6.0	5.3

*From Spache, 1978.

Rogers (1970) applied the Spache and Fry formulas to 13 fourth-grade basal readers, and he found two distinct ranges of difficulty for these books. The more difficult readers had a range of readability estimates from 3.5 to 6.0, with a mean of 4.5. The easier group of books ranged from 2.8 to 3.9, with a mean of 3.2. It is evident from these studies that the reading levels that publishers assign to their books cannot be relied upon to represent the actual overall readability levels of those books.

There is also a great deal of variability among different diagnostic reading tests when it comes to the actual readability levels of passages that are supposedly at the same grade level. Hood (unpublished) found that this variability is largely due to the choice of readability measures used in constructing and grading the test passages.

Using Readability Measures to Match Students with Books

Suppose on a given diagnostic reading test the reading achievement level of Student A is determined to be early third grade. And suppose that using a given readability measure, Book B is found to be early third-grade level in difficulty. Can we be certain that Student A can be instructed effectively using Book B? Given the fact that the reading ability of a student is variously estimated by different reading tests and the reading difficulty of a book is variously estimated by different readability measures, the answer is no.

One way of rising above the confusion is to select your favorite readability measure and apply it to the test passages of your favorite diagnostic reading test. Then apply this same readability measure to several passages from various books in the series that is being considered for a student's reading instruction. This way, the test passages and

book passages will be truly comparable, having been subjected to the same standard of measurement. Thus, if the student you are evaluating obtains an instructional reading level of 3-1 (early third grade) on the reading test (e.g., he reads the passage that has a 3-1 readability level with 95 percent oral reading accuracy), then you can be fairly confident that the most appropriate book for his reading instruction is the one with a readability level of 3-1. The most suitable reading tests for this purpose would be criterion-referenced tests containing oral reading passages. (e.g., the Standard Reading Inventory). The Gray and the Gilmore are two standardized tests that could also be used in this approach, since both of them, like the criterion-referenced tests, were designed to be representative of the material found in various basal readers used for reading instruction (Hood, unpublished, p. 39).

Constructing and Informal Reading Inventory

More involved and time-consuming than the approach described above is that of constructing an informal reading inventory (IRI), in which the test passages the student reads come directly from the books that are being considered for his reading instruction. Thus, if the student reaches his instructional reading level (to be defined in the next section) on a given test passage drawn from a particular book, then that book is probably the best choice of instructional materials for that student. (This depends on whether the test passage is truly representative of the book from which it was taken in terms of readability level.)

For example, if the student being evaluated is in a class using the Houghton-Mifflin basal reader series, he should be tested on

representative passages from several books in that series. Then, his performance on this Houghton-Mifflin-based IRI should indicate exactly which Houghton-Mifflin book he should be placed in for reading instruction. In this way, the student's reading test score becomes a very useful bit of information that translates directly into a practical recommendation for the teacher.

In describing how to construct an IRI, the author will rely mainly on information she learned during her brief training in the Reading Clinic at the University of Iowa. This information deals specifically with IRI's for elementary school children, or for older students whose reading level is assumed to be sixth grade or below.

First, the examiner must obtain a copy of each book in the series that is being considered for the student's reading instruction. Preferably, the series of books should be a well-graded system which covers at least three grade levels or, better yet, covers all levels from preprimer through sixth grade. Selected passages should sound complete; they should make sense in the context of the entire story. They should also be passages about which several good comprehension questions can be asked. Recommendations as to length of the passages vary from one authority to another. Since most widely-used readability formulas are designed for 100-word passages, it is probably sensible to have all passages be at least 100 words, although at the preprimer level 30 words may be enough, and at the primer and first reader levels, 60 to 70 words may be sufficient. In books from grades 2 to 6, passages should be 150 to 200 words.

Two selections should be chosen from each book. At the preprimer, primer, and first reader levels, one passage should come from near the beginning and one from near the end. In books 2-1, 2-2, 3-1, 3-2, 4,

5, and 6, two passages should be selected from near the center of the book--one for oral reading and one for silent reading.

Since publishers' estimates of readability are not accurate enough for the purpose of matching a student with appropriate materials, it is necessary to determine the readability levels of all passages chosen. The Botel Readability formula should be used for levels preprimer through third grade, and the Dale-Chall Readability formula (revised) should be used for fourth grade and beyond. This may mean trying out several passages before you find one that fits the correct difficulty level. It is also important that each passage be representative of the overall readability of the book or section of the book from which it was selected. This means that the overall readability level of each book should be estimated by applying the readability formula to several passages throughout the book and computing the average of the levels obtained--a time-consuming task but one that will make your IRI much more valid and useful. Some publishers provide a series of selections to be used as oral reading inventories. These selections are presumably representative of the readers from which they are selected, but it would be prudent to substantiate this presumption by means of readability estimates.

The passages should be placed in hierarchical order and no levels between the lowest and the highest should be omitted. The levels to be included are: preprimer (PP), primer (P or 1-1), 1-2, 2-1, 2-2, 3-1, 3-2, 4, 5, and 6. The range of passages administered to any given child will probably be smaller than this (e.g. PP to 3-1), but every level between his basal and ceiling levels should be administered.

Comprehension questions should be constructed carefully. Questions

requiring a "yes" or "no" answer or pure recall of words should be avoided. Types of questions should vary and should include a balance of factual questions, inferential questions, and questions dealing with vocabulary. The number of questions should increase with the reading level of the passage, from about five questions at the primer level to about eight questions for grades 3 to 6. (Further discussion of the issue of comprehension will follow in a later section.)

The child may read selections directly from the reader or the material may be clearly typed or cut out of a consumable source and mounted in a loose-leaf notebook, with only one passage on each page. The examiner should have a typed, double-spaced copy of the selections to allow for easy scoring, and the comprehension questions should be listed at the bottom of the page.

Selecting Scoring Criteria for Informal Reading Inventories

The validity of an IRI, as with any criterion-referenced reading test, is partly determined by the comparability of its content to the content of instructional materials. Its validity also depends on the appropriateness of the criteria used to determine satisfactory reading achievement. Unfortunately, there are no unanimously accepted criteria; different authorities offer differing recommendations as to what criteria should be used in scoring IRI's and determining children's reading levels.

Which Oral Reading Errors Are Significant?

There is a great deal of disagreement over which oral reading errors should be counted as errors and which should be overlooked as insignificant. A comparison of several standardized reading tests (both normed and criterion-referenced) shows considerable variation in the counting of errors. Hood (unpublished) examined which oral reading errors are considered

scorable errors on the following tests: Classroom Reading Inventory, Diagnostic Reading Scales, Gates-McKillop Reading Diatnostic Tests, Gilmore Oral Reading Test, Gray Oral Reading Tests, and Standard Reading Inventory. She found agreement among all six tests that unknown words (which may be pronounced for the child by the examiner) are to be counted as errors, as well as omissions, insertions (additions), repetitions, mispronunciations, and substitutions of whole words, parts of words, or nonsense words.

There is little agreement among the tests on which other errors should be counted. Only the Gates-McKillop specifically mentions that substitutions of contractions for text words should be considered errors. The Gilmore and the Standard Reading Inventory count punctuation errors (though their definitions of a punctuation error differ); none of the other tests mentions punctuation errors. Only the Gilmore counts hesitation errors, whereas the Diagnostic Reading Scales, the Gates-McKillop, and the Standard Reading Inventory specifically state that hesitations should not be counted. Self-corrections are to be counted as errors in the Gilmore and the Standard Reading Inventory, but not in the Gray Oral or the Diagnostic Reading Scales, and they are not mentioned in the scoring guidelines of the Gates-McKillop or the Classroom Reading Inventory. Thus, the errors which should be counted are not always clearly specified in the scoring manuals of these tests, nor is there complete agreement on the scoring procedures that are specifically described.

Packman (1972) studied the relative importance of different kinds of oral reading errors. She investigated the effect of selected types of errors on the comprehension scores obtained on the test passages

in which the errors were made. She reported that the number of pronunciation errors (words supplied by the examiner), mispronunciations, omissions, substitutions, and repetitions all increased as the level of comprehension decreased. In contrast, there was no increase in addition errors and punctuation errors as the level of comprehension declined.

McCracken (1967) recommends that the following be counted as oral reading errors in scoring IRI's:

(1) Repetition -- repeating a word or phrase;

(2) Substitution -- saying a different word instead of the text word;

(3) Addition -- inserting a word or words or adding an affix;

(4) Pronunciation -- the examiner pronounces an unknown word for the child;

(5) Omission -- omitting a word, phrase, or affix;

(6) Mispronunciation -- saying a word in a manner which is definitely incorrect and not a result of defective speech or colloquial pronunciation;

(7) Punctuation -- phrasing in which the punctuation is definitely misread or added. Ignoring punctuation is not an error.

This psychologist agrees with McCracken's list of scorable oral reading errors, except for the category called "pronunciation". It seems that if the examiner pronounces words for the child, the examiner's word recognition ability is being measured along with the child's. This practice is likely to falsely inflate comprehension scores, especially for the child who habitually hesitates on unknown words, waiting for the examiner to tell him what they are. It would seem preferable to encourage the child to try every word, or even to skip a word if he cannot bring himself to try it. Then, depending on the child's response, the attempt

can be scored as correct, as a mispronunciation, as an omission, or whatever it happens to be.

McCracken (1967) offers these guidelines for questionable or complicated scoring situations:

(1) Count only one error at any one place in the reading. A student often makes more than one type of error at one point in the passage. For example, a student may omit a difficult word, reread the phrase containing the word (repetition), mispronounce the omitted word, then reread again (another repetition) and pronounce the word correctly. All of this would be counted as one error.

(2) Count as one error if a student corrects an error, with or without repeating other words.

(3) Count as one error the omission or addition of two or more consecutive words.

(4) Count as one error if the child makes a second error caused by his forcing grammatical agreement with his first error.

(5) Count as one error the mispronouncing of a proper name or difficult word if the word appears more than once in a passage and is mispronounced more than once. Also count as one error if a proper name has two or more words in it and both are mispronounced. Count errors on simple words (e.g., "a", "the") each time they occur.

These guidelines appear to make error scoring more consistent and precise, and they are recommended for use by those administering IRI's.

Three Levels of Reading Proficiency

Administration of an IRI can determine three levels of reading proficiency for a student: (1) the independent level--the level at which the child is ready to function independently; (2) the instructional level--the level at which he can now profit from instruction; and (3)

the frustration level--the level where he reaches complete frustration with the material.

The original standards for the independent, instructional, and frustration levels were suggested by Betts (1946), and these standards or slight variations of them are still accepted by many advocates of the informal reading inventory (Johnson and Kress, 1968; McCracken, 1967). The traditional criteria for the three reading levels are as follows:

	Word Recognition	Comprehension
Independent level	99% +	90% +
Instructional level	95% +	75% +
Frustration level	90% -	50% -

Johnson and Kress (1968) describe in detail the criteria to be applied in determining each of these reading levels for a child.

Independent level. At the independent level, they say, the child can function on his own and do a virtually perfect job of handling the material. His silent reading should be free from observable symptoms of difficulty such as finger pointing, vocalization, lip movement, and other signs of general tension while reading. Oral reading should be done in a rhythmical fashion and a conversational tone. As he reads, the child should make no more than one error in 100 running words (99 percent word recognition accuracy). He should be able to answer at least 90 percent of the comprehension questions correctly.

Instructional level. The instructional level is that at which the child should be and can profitably be instructed. Again the child should be free from observable signs of difficulty and tension, and he should be able to read rhythmically and in a conversational tone. He should be able to achieve 95 percent word recognition accuracy when reading a passage he has never seen before, and his accuracy should improve upon

rereading. In terms of comprehension, he should be able to attain a 75 percent level of understanding of the material. Johnson and Kress (1968) state that when a child receives instruction at this level, "...he in all probability will be able to reach, with teacher help, the same high levels of performance as were indicated as criteria for the independent level. In general, one should strive in instruction to have the child handling the material independently by the time the lesson is completed. If he begins the lesson with less adequacy than indicated in these criteria, there is very little likelihood that he will overcome all of his problems."

Frustration level. This is the point at which the child becomes completely unable to handle the material. The child may show observable signs of tension and difficulty with the material. His word recognition accuracy will be 90 percent or less, and his comprehension will be 50 percent or less. Johnson and Kress (1968) feel that knowing this level may serve two purposes for the child's teacher. It will tell him what levels of material to avoid for this child's work. It may also give him some indication of the rate at which the child might be able to progress when he is taught at his proper instructional level. If a child is ready for instruction at one level and completely frustrated at the next, there is evidence that he has many problems to be overcome through instruction at the appropriate level before he will be ready for the next level, and progress is likely to be slow. On the other hand, if there is considerable spread between the child's instructional and frustration levels, there is a better chance for fairly rapid progress, because his problems at the instructional level are probably not as severe or complex.

Validity of the Traditional Standards for Reading Levels

A number of authors such as Pennock (1975) and Powell (1970)

have suggested that the criteria for determining reading levels may be too stringent and too hard on students. In contrast, Johnson and Kress (1968) feel that all too often the criteria used for determination of reading levels are too low, and they support the traditional criteria as being in line with their experience. They state, "Experience has shown that when there is too much to be accomplished through instruction, the child does not perform adequately in terms of profiting from instruction and retaining those things which are taught." Research evidence will be presented below that attempts to validate the traditional standards for determining levels of reading proficiency that have been outlined above.

Hood (unpublished) discussed an unpublished study by Moore (1972) which investigated gains in reading achievement made by children instructed at their independent, instructional, or frustration levels (defined by the criteria recommended by McCracken (1967). The error-count included omissions, insertions, substitutions (including mispronunciations), and examiner-provided words, but it did not include repetitions. The 21 students in her study were third-graders who were considered average achievers by their teachers. The children were divided into three groups, each of which received six weeks of instruction at each of the three difficulty levels so that the children served as their own controls. The order of instruction was partially counterbalanced to control for order effects. The gains in reading achievement were greatest during the periods when instruction was given at the instructional level, and gains were greater during instruction at the frustration level than at the independent level. The statistical level of confidence in this study is low. However, it provides some evidence in favor of the

validity of the instructional level, which is supposed to be the level at which the student can profit the most from instruction.

Evidence for the validity of the criterion of ten percent word recognition errors for the frustration level comes from a study by Eckwall (1974). Since outward signs of tension and anxiety are commonly present when a student is reading in material at his frustration level, Eckwall sought to measure this anxiety more objectively by using a polygraph to measure physiological frustration in the subjects. At the frustration level stipulated by the polygraph examiner, the percentages of oral reading errors and comprehension errors were computed. There was no significant difference between the commonly accepted criterion of ten percent oral reading errors (90 percent accuracy) and polygraph-measured frustration level. This comparability between the traditional standard for frustration level and an objective measure of frustration appears to provide some evidence that the traditional standard is a valid one.

The error count in the Eckwall (1974) study included uncorrected omissions, insertions, substitutions (including mispronunciations), requests for examiner aid, and repetitions. Eckwall also computed error percentages when repetitions were not included in the error count, and he found that subjects were reaching frustration according to polygraph when they were making only 7.65 percent oral reading errors on the average (92.35 percent oral reading accuracy). Eckwall noted that, in using the 90 percent oral reading accuracy criterion is used, "...if one does not count repetitions as errors, then a student is quite likely to become physiologically frustrated before the examiner has recorded enough errors to actually designate the frustration level." Eckwall uses these results to argue that authors such as Pennock (1975) and

Powell (1970) are wrong in urging that criteria for reading levels be made less stringent (an example of which would be to not include repetition errors in the error count). He feels that to lower the standards would very often place children in material that is too difficult for instructional purposes.

The two studies described above provide some evidence for the validity of the instructional and frustration levels as they are defined by the traditional standards for oral reading errors. Although more research certainly is needed, it seems safest at this point to accept the traditional oral reading error criteria for determining levels of reading proficiency and to use these standards in interpreting children's performance on IRI's.

Difficulties in Measuring Comprehension on IRI's

What about the traditional comprehension criteria for establishing reading levels? These criteria are comprehension levels of 90 percent or above at the independent level, 75 percent or above at the instructional level. Are these criteria appropriate?

Hood (unpublished) questions the common practice of preparing comprehension questions to accompany reading selections and then considering the percent of questions correctly answered as an important part of the criteria for choices of instructional materials. Hood argues, "The oral reading inventory is appropriately used for instructional placement only when a child is unable to recode printed messages which he can decode (understand) if they are read to him. In this situation the determination of the comprehension level as a basis for instructional placement would appear to be irrelevant. A child who is diagnosed as

having word recognition difficulties would be a child who, by definition, would be able to comprehend a selection that is appropriate for his age level, if only he could recode the words. A child who could recode the words and then could not comprehend the message would be diagnosed as having a comprehension difficulty instead. The oral reading inventory is not appropriate for determining instructional placements for children who have comprehension difficulties, whether or not questions are prepared to accompany the passages."

Hood (unpublished) feels that the appropriate way to measure comprehension is through silent reading, rather than oral reading. However, she even questions the assumption that the comprehension questions following silent reading passages might be useful for determining reading levels for children with comprehension difficulties. Hood states, "This assumption is questionable because of the difficulty of preparing fair questions over test passages without employing standardization procedures which include extensive item tryouts. As previously argued in this discussion, it is possible to prepare easy questions over hard selections and hard questions over easy selections. A teacher could not know whether the questions he or she had prepared were of the appropriate difficulty level for the average child who might be able to recode a selection unless a properly controlled experiment had been conducted to evaluate them."

Pikulski (1974) also discusses the difficulties in measuring reading comprehension, and he notes that the problem is not unique to informal reading evaluation. One specific problem he notes is the difficulty of ensuring that the questions asked are reading-dependent questions--that is, that they can be answered only with reference to the information

contained in the reading selection. Insisting that all questions be reading-dependent would largely eliminate vocabulary questions like: "What is a beach? What is a ticket?" It would also eliminate questions that ask for more information than the selection contains and depend on the breadth of background experience of the student. Although vocabulary and broad experience are factors which undoubtedly contribute to good reading performance, they do not help in determining how well a student understood a particular passage, and they may give an unfair advantage to students of certain socioeconomic groups or cultural backgrounds.

Hood (unpublished) is also concerned that questions be reading-dependent. She points out that some questions can be answered by most children without having read the passage. Hood gives as an example the question, "Why did she stop at the corner?" (question number 5 for passage 1A in Spache's test). It is likely that many first graders might guess the correct answer without reading the passage at all. Comprehension scores based on such questions would be meaningless.

Hood (unpublished) recommends that IRI's be used specifically for children who are experiencing word recognition difficulties, and she appears to feel that the comprehension criteria for IRI's should be ignored. For children with comprehension difficulties, she feels, the usual type of IRI is not appropriate. For these children, she suggests devising an informal comprehension inventory, drawn directly from published instructional materials designed to improve children's comprehension skills. Samples of silent reading comprehension exercises at increasing levels of difficulty could be taken directly from the comprehension materials the student is likely to be using for instruction.

His performance on these exercises would determine at what level his instruction in comprehension skills should begin. The lowest level at which the child began to show some difficulties and to require some guidance from the teacher would be the best level at which to begin for instructional purposes.

Of course, there will still be many school psychologists and others using IRI's who feel that the importance of reading comprehension makes it essential to include questions at the end of each reading selection. One solution would be to include comprehension questions in the testing procedure, as a rough indication of the child's level of understanding, but simply not use the obtained comprehension scores as part of the criteria in determining levels of reading proficiency. This would eliminate the need for silently-read passages; these passages could be read orally instead, thereby providing a larger sample of the student's oral reading ability.

For those who are determined to write good comprehension questions, in spite of the difficulties inherent in such a task, Valmong (1972) has some excellent suggestions. The kinds of questions to be included in an IRI, according to Valmont, are main idea questions, detail questions, inferences, drawing conclusions, organization questions, cause and effect questions, and vocabulary questions. He further suggests several important guidelines to follow in creating these questions, some of which are listed below:

(1) Questions should be in the order in which the information to which they refer occurred in the passage.

(2) Main idea questions, when included, should be first.

(3) Ask the most important questions possible.

(4) Make sure a later question is not answered by an earlier one.

(5) Except for vocabulary questions, a question should only be answerable by someone who has read the passage.

(6) The question should not require outside experience to be answered correctly.

(7) Pictures should not aid the child in answering a question.

(8) Questions should be kept short, simply-worded, and relevant.

(9) Good questions often start with who, what, when, where, how, or why.

(10) Do not state questions in a negative manner.

(11) Each question must have only one correct answer.

(12) Do not use questions that can be answered "yes" or "no".

Yet another solution to the problem of creating good comprehension questions is to use existing criterion-referenced diagnostic reading tests, such as the Standard Reading Inventory or the Classroom Reading Inventory, in place of an IRI. As discussed earlier in the paper, if one applies the same readability formula to both the test passages and the reading materials to be used in instruction, a student's performance on one of these tests can be translated directly into an appropriate choice of instructional materials for him. For instance, if the student reaches his instructional level (95 percent oral reading accuracy) on the Standard Reading Inventory paragraph that has a 3-1 readability level according to the Botel readability measure, then he can be placed for reading instruction in the Houghton-Mifflin (or any other) reader that also has an overall readability level of 3-1 according to the Botel method.

The advantage of this approach is that someone else, in this case McCracken (1966), has designed your reading inventory for you and,

presumably, his comprehension questions are very carefully constructed, with a properly controlled experiment conducted to evaluate them. The validity of this presumption should be checked, however.

After Instructional Materials Are Selected--Then What?

All the diagnostic reading tests discussed in this paper utilize a quantitative count of oral reading errors as a criterion for determining the instructional reading level of a student. This information can lead to the selection of the most appropriate materials to use in instructing that student. But once the best materials are chosen, how does one determine the best procedures to use for improving that student's reading ability?

What is needed is a method of analyzing the student's oral reading errors in more detail, to determine more specifically what aspects of reading are giving him trouble and which of those trouble spots require the greatest or the most immediate attention. The purpose of such an analysis would be to determine what specific procedures should be followed in instructing this student at his instructional level.

Reading Miscue Inventory

It is for this purpose that the Reading Miscue Inventory Procedure for Diagnosis and Evaluation has been prepared. (Burke and Goodman, 1972). The Reading Miscue Inventory does not provide guidelines for instructional placement with reference to readability levels nor does it suggest an appropriate criterion for determining acceptable oral reading accuracy. Instead it offers a procedure for the qualitative analysis of oral reading miscues (errors) as a basis for planning reading strategy lessons--a personalized reading program for a particular child.

The rational for the Reading Miscue Inventory, as discussed by Yetta Goodman (1972), is that the extent to which meaning is disrupted by errors (called miscues) is more significant than a simple error count as an indication of how much sense a student is making out of the printed words. (For example, cap and camp are often miscues which distort the meaning of the sentence or the passage, and, thus, interfere significantly with the student's comprehension.) Miscues which are semantically similar to the printed words (e.g., cap - hat) are less likely to distort meaning and, therefore, do not interfere as much with the student's comprehension of the material. In this sense, graphically similar miscues are more significant because they indicate greater reading difficulties than do semantically similar miscues.

In gathering information for the Reading Miscue Inventory, Burke and Goodman (1972) recommend that a child read an entire selection or several selections lasting from 15 to 20 minutes, and that the selection be difficult enough to provide a minimum of 25 miscues for the analysis. Hood (unpublished) feels that the selections should be at the child's instructional level in terms of readability, since it has been suggested that the pattern of oral reading errors may change depending on the child's difficulty with the material being read (McCracken, 1967). As the child reads, the examiner marks errors as he would for an IRI. Afterward the examiner lists all the miscues and analyzes them by asking nine questions with regard to each miscue:

(1) Dialect: Is a dialect variation involved in the miscue? (yes or no)

(2) Intonation: Is a shift in intonation involved in the miscue? (yes or no)

(3) Graphic similarity: How much does the miscue look like what was expected? (high, partial, or no similarity)

(4) Sound similarity: How much does the miscue sound like what was expected? (high, partial, or no similarity)

(5) Grammatical function: Is the grammatical function of the miscue the same as that of the text word? (yes, can't determine, or no)

(6) Correction: Is the miscue corrected? (corrected, correction unsuccessful or abandoned, or no correction attempted)

(7) Grammatical acceptability: Does the miscue occur in a structure which is grammatically acceptable? (yes, acceptable in sentence but not in relation to preceding or following sentences, or no)

(8) Semantic acceptability: Does the miscue occur in a structure which is semantically acceptable? (yes, acceptable in sentence but not in relation to preceding or following sentences, or no)

(9) Meaning change: Does the miscue result in a change in meaning? (extensive, minimal, or no meaning change)

Each of the reader's miscues is categorized according to this scheme. For example, if the sentence, "I looked up when I heard the bell," were read as, "I looked up when I had the bell," the substitution of "had" for "heard" would not be considered the result of dialect, there was no shift in intonation, there was a strong graphic similarity, a moderate amount of sound similarity, the two words serve the same grammatical function, it was grammatically acceptable but not semantically acceptable, and there was an extensive change in meaning. By analyzing each of the child's miscues in this manner, one can discover his word recognition strategies and evaluate his strengths and weaknesses. From such an evaluation, lessons which use the child's strengths to overcome his weaknesses can be devised and used.

Conclusion

Undoubtedly many school psychologists are quite familiar with the issues discussed in this paper, and perhaps many of them are already employing some of the procedures presented here. Those of you who are already using readability formulas, informal reading inventories, and other such tools in determining the instructional needs of the students you evaluate are to be congratulated. For those of you who have been reporting reading test scores and leaving it at that, it is hoped that you have found some tools and procedures in this paper that you intend to try out in the future. By utilizing some of these techniques, you will surely be increasing the educational relevance of your evaluations, the usefulness of your recommendations, and the reading progress made by the children you evaluate.

REFERENCES

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

Bond, Guy L. & Tinker, Miles A. Reading Difficulties: Their Diagnosis and Correction. (2nd ed.) New York: Appleton-Century-Crofts, 1967.

Botel, Morton, Predicting Readability Levels. Chicago: Follett, 1962.

Botel, Morton. How to Teach Reading. (4th ed.) Chicago: Follett, 1968.

Burke, Carolyn L. & Goodman, Yetta M. Reading Miscue Inventory Procedure for Diagnosis and Evaluation. New York: MacMillan, 1972.

Cohen, S. A. "Studies of Visual Perception and Reading in Disadvantaged children". Journal of Learning Disabilities, 1969, 2, 498-507.

Dale, Edgar & Chall, Jeanne S. "A Formula for Predicting Readability". Educational Research Bulletin, Jan. 21 and Febr. 17, 1948, 27, 11-20, 37-54.

Edkwall, Eldon E. "Should Repetitions Be Counted as Errors?" The Reading Teacher, 1974, 27, 365-367.

Farr, Roger. Reading: What Can Be Measured? Newark, Delaware: International Reading Association, 1969.

Fisher, Dennis R. & Frankfurter, Anthony. "Normal and Disabled Readers Can Locate and Identify Letters: Where's the Perceptual Deficit?" Journal of Reading Behavior, 1977, 9, 31-43.

Flesch, Rudolf F. "A New Readability Yardstick". Journal of Applied Psychology, 1948, 32, 221-233.

Fry, Edward B. "A Readability Formula That Saves Time". Journal of Reading, 1968, 11, 513-516, 575-578.

Goodman, Yetta M. "Reading Diagnosis--Qualitative or Quantitative". The Reading Teacher, 1972, 26, 32-37.

Hammill, Donald D. "Training Visual Perceptual Processes". Journal of Learning Disabilities, 1972, 5, 39-46.

Harris, Albert J. How to Increase Reading Ability. (5th ed.) New York: David McKay Co., 1970.

Harris, Albert J. "A Comparison of Formulas for Measuring Degree of Reading Disability". In Robert E. Leibert (Ed.), Diagnostic Viewpoints in Reading. Newark, Delaware: International Reading Association, 1971.

Hood, Joyce. "Matching the Readability of Instructional Materials to the Reading Ability of the Child". Unpublished manuscript, undated. The University of Iowa.
(Copies available from Joyce Hood, W313 East Hall, College of Education, University of Iowa, Iowa City, Iowa 52242.)

Johnson, Marjorie Seddon & Kress, Roy A. "Individual Reading Inventories". In Leo M. Schell & Paul C. Burns (Eds.), Remedial Reading: An Anthology of Sources. Boston: Allyn & Bacon, Inc., 1968.

Klare, George R. The Measurement of Readability. Ames, Iowa: Iowa State University Press, 1963.

Klare, George R. "Assessing Readability". Reading Research Quarterly, 1974-75, 10, 62-102.

Kolson, C. J. & Kaluger, G. Clinical Aspects of Remedial Reading. Springfield, Illinois: Charles C. Thomas, 1963.

Larsen, Stephen C. & Hammill, Donald D. "The Relationship of Selected Visual Perceptual Abilities to School Learning". Journal of Special Education, 1975, 9, 281-291.

Lewerenz, Alfred S. "Selection of Reading Materials by Pupil Ability and Interest. Elementary English Review, 1939, 16, 151-156.

Maginnis, George H. "The Readability Graph and Informal Reading Inventories. The Reading Teacher, 1969, 22, 516-519.

Mann, L. "Perceptual Training: Misdirections and Redirections". American Journal of Orthopsychiatry, 1970, 40, 30-38.

McCall, William A. & Crabbs, Lelah M. Standard Test Lessons in Reading. New York: Bureau of Publications, Teachers College, Columbia University, (3rd edition), 1961.

McCracken, Robert A. "The Informal Reading Inventory as a Means of Improving Instruction". In Thomas C. Barrett (Ed.), The evaluation of children's reading achievement. Perspectives in Reading Series, No. 8. Newark, Delaware: International Reading Association, 1967, 79-97.

Moore, Sandra Kay. "A Comparison of the Progress of Third Grade Children When Taught in Materials of Varying Levels of Difficulty". Unpublished doctoral dissertation, University of Iowa, 1972.

Packman, Linda. "Selected Oral Reading Errors and Levels of Reading Comprehension". In Howard A. Klein, (Ed.), The Quest for Competency in Teaching Reading. Newark, Delaware: International Reading Association, 1972.

Pennock, Clifford. "IRI--Dogma or Data". The Reading Teacher, 1975, 28, 782-783.

Pikulski, John. "A Critical Review: Informal Reading Inventories". The Reading Teacher, 1974, 28, 141-151.

Powell, William R. "Reappraising the Criteria for Interpreting Informal Reading Inventories". In D. DeBoer (Ed.), Reading diagnosis and evaluation. Newark, Delaware: International Reading Association, 1970.

Rogers, A. R. "Comparing the Difficulty of Basal Readers". Journal of the New England Reading Association, 1970, 5, 35-37.

Spache, George D. "A New Readability Formula for Primary Grade Reading Materials". Elementary School Journal, 1953, 53, 410-413.

Spache, George D. Good Reading for Poor Readers. Champaign, Illinois: Garrard, 1966.

Spache, George D. Good Reading for Poor Readers. (9th edition) Champaign, Illinois: Garrard Press, 1974.

Spache, George D. Investigating the Issues of Reading Disabilities. Boston: Allyn & Bacon, Inc., 1976.

Stone, Clarence R. "Measuring Difficulty of Primary Reading Material: A Constructive Criticism of Spache's Measure". Elementary School Journal, 1956, 57, 36-41.

Valmont, William J. "Creating Questions for Informal Reading Inventories". The Reading Teacher, 1972, 25, 509-512.

Young, Beverly S. "A Simple Formula for Predicting Reading Potential". The Reading Teacher, 1976, 29, 659-661.

READING TEST REFERENCES

BOTEL READING INVENTORY. Morton Botel. Chicago: Follett Educational Corporation, 1970.

CLASSROOM READING INVENTORY. Nicholas J. Silvaroli. Dubuque, Iowa: Wm. C. Brown Company Publishers, 1969.

DIAGNOSTIC READING SCALES. George D. Spache. Monterey, California: McGraw Hill, Inc., 1963.

DURRELL ANALYSIS OF READING DIFFICULTY, New Edition. Donald D. Durrell. New York: Harcourt, Brace & World, Inc., 1955.

GATES-MCKILLOP READING DIAGNOSTIC TESTS. Arthur I. Gates and Anne S. McKillop. New York: Teachers College Press, 1962.

GILMORE ORAL READING TEST. John V. and Eunice C. Gilmore. New York: Harcourt, Brace & World, Inc., 1968.

GRAY ORAL READING TESTS. William S. Gray, ec. by Helen M. Robinson. Indianapolis: The Bobbs-Merrill Company, Inc., 1967.

STANDARD READING INVENTORY. Robert A. McCracken. Klamath, Oregon: Klamath Printing Company, 1966.