DOCUMENT RESUME

ED 064 655 24 CG 007 503

AUTHOR Calfee, Robert C.; Venezky, Richard L.
TITLE Component Skills in Beginning Reading.
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
REPORT NO TR-60
BUREAU NO BR-5-0216
PUB DATE Jul 68
CONTRACT OEC-5-10-154
NOTE 19p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Cognitive Development; Intelligence; *Language Development; Performance Factors; *Reading; Reading Readiness; Reading Skills; *Reading Tests; *Skill Development

ABSTRACT

Reading research in which different methods or materials have been compared has proven inconclusive. This paper is restricted to beginning reading, defined as the acquisition of letter-sound decoding ability, and raises the question: what skills are required by current tests? Available reading readiness and achievement tests consist of batteries of subtests, each of which is designed to measure a component skill necessary in reading. However, high intercorrelations between the subtests indicate either that separable skills are not being measured, or that skills develop at the same rate in most children. However, the makeup of the items in the tests is such that ability to follow instructions and general language competence are common factors which enter significantly into performance on all subtests. The experience of psychologists in constructing tests to identify separable skills in language and intelligence indicates that this task is possible but difficult. Current tests are suitable for prediction of reading performance, but tests that evaluate separable skills are urgently needed for further research on the development of the reading process, as well as diagnosis. Examples are presented for articulation and phonetic discrimination. (Author)

COMPONENT SKILLS IN BEGINNING READING

WISCONSIN RESEARCH AND DEVELOPMENT

CENTER FOR

OGNITIVE LEARNING

RESEARCH
COGNITIVE
LEARNING
DEVELOPMENT

Technical Report No. 60

COMPONENT SKILLS IN BEGINNING READING

Robert C. Calfee and Richard L. Venezky

Report from the Project on Language Concepts and Cognitive Skills
Related to the Acquisition of Literacy

R. C. Calfee and R. L. Venezky, Principal Investigators

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

July 1968

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare, under the provisions of the Cooperative Research Program.

Center No. C-03 / Contract OE 5-10-154

PREFACE

A goal of this Center is to create knowledge and theory which can be effectively utilized in the construction of instructional systems for the schools of tommorow. Some researchers prefer to begin by constructing new instructional materials immediately, while others prefer to begin by studying the fundamental processes presumed to be required for the mastery of such instructional materials. Regardless of the particular approach, a Research and Development Center should ideally provide an atmosphere within which scholars with different techniques and areas of competence but with common interests can form effective research teams. Such a team is represented in this project, with Professor Calfee from Psychology, and Professor Venezky from English and Computer Sciences.

While their ultimate goal is the construction of reading materials which will optimize reading acquisition, these researchers are presently attempting to gain a better understanding of the fundamental independent cognitive skills related to the reading process. This report contains the rationale for their approach and the results of their analyses of existing tests of component reading skills. Significantly, the authors conclude that existing diagnostic tests do not measure independent skills. However, the authors express confidence that sensitive tests can be developed for measuring the independent cognitive skills related to reading, and, in the process for prescribing remedial treatment for those children lacking these prerequisite skills.

Harold J. Fletcher
Director of Program 1

CONTENTS

LIST OF TABLES

ABSTRACT

Reading research in which different methods or materials have been compared has proven inconclusive. This paper is restricted to beginning reading, defined as the acquisition of letter-sound decoding ability, and raises the question: what skills are required by current tests? Available reading readiness and achievement tests consist of batteries of subtests, each of which is designed to measure a component skill necessary in reading. However, high intercorrelations between the subtests indicate either that separable skills are not being measured, or that skills develop at the same rate in most children. However, the makeup of the items in the tests is such that ability to follow instructions and general language competence are common factors which enter significantly into performance on all subtests. The experience of psychologists in constructing tests to identify separable skills in language and intelligence indicates that this task is possible but difficult. Current tests are suitable for prediction of reading performance, but tests that evaluate separable skills are urgently needed for further research on the development of the reading process, as well as diagnosis. Examples are presented for articulation and phonetic discrimination.

COMPONENT SKILLS IN BEGINNING READING

INTRODUCTION

Suppose that the time and money available for further research on improvements in reading instruction were limited. Given but a year or two of support and tight limitations on the budget, what research would have highest priority? Russell and Fea (1965) point out in their review of reading research that no other area of the curriculum has garnered such a huge pile of reports. Nevertheless, despite thousands of studies over the past 50 years, there is no clear evidence of improvements in reading instruction or significant changes in instructional technique.

The majority of experiments on reading have explored the relative efficiency of various methods or materials. One finds comparisons between phonic programs and whole-word approaches, between ITA and TO, studies of the effects of different grouping practices, of stressing comprehension or drill, and arguments about the effectiveness of visual or auditory presentation. Type font size, the kind of pictures accompanying the reading text, the style and content of the vocabulary, and the length and placement of sentences have been examined. Of the many techniques that have been tried, (a) most seem to work with most children, but all fail with many children; (b) there appears to be no best method; and (c) the efforts of the teacher appear to override in importance the effects of variation in methods or materials—or so goes the folklore.

Bond and Dykstra (1967) in the report of the Coordinating Center for the Cooperative Research Program in First-Grade Reading Instruction present data from 27 research projects. From this extensive report, the conclusions most pertinent to the effectiveness of different methods were (a) various innovative methods, whether phonic, linguistic, orthographical, language experience, or what have you, produced reading achievement scores at the end of the first grade that were slightly higher than basal reader methods; (b) these differences were generally small and were not consistently observed by all researchers in all school systems; and (c) there was no evidence of differential effectiveness (i.e., it was not true that some methods worked better with low IQ students and others with high IQ students). It was further concluded that reading achievement must be determined by many factors of equal or greater importance than those examined in the report (i.e., other than readiness, IQ, method/material variation, teacher experience, and community background, etc.). Although attention was directed to the need for more adequate teacher training, none of the teacher variables measured (sex, age, education, certification, experience, attitude toward teaching, and rated effectiveness) bore any substantial relation to reading performance.

A Hawthorne or novelty effect may have led to the slight superiority of the several innovative methods. Chall (1967) has pointed to several sources of novelty—fresh books and supplementary materials, special training for the teacher, and the knowledge by students and parents that they were being treated differently.

COMPONENTS OF THE READING PROCESS

In reviews and research reports, one frequently finds reference to the reading process. For example, Russell and Fea (1965) speak of "the reading act as consisting of two components, (a) identifying the symbol, and (b) obtaining meaning from the identified symbol." Levin (1966) refers to one skill as decoding the written language into its spoken form, and a second skill as the use of this decoding ability for comprehension. Other authors have expressed the distinction most succinctly as learning to read and reading to learn.

If he is to become literate, the child must somehow acquire the ability to decode or translate written material to that form of the spoken

language with which he is already familiar. This skill may assume different forms over time. A beginning reader, haltingly translating single words or phrases, almost certainly uses different psychological operations than those available to the accomplished reader who can skim a paragraph or a page in a matter of moments.

In this paper, we will be concerned primarily with the acquisition of a rudimentary decoding ability. If the ability to translate from letters to sounds is considered a complex skill, then the individual must have at his disposal certain more basic skills which are augmented and integrated during the acquisition of the new, more complex skill. It is natural to ask, what are the component skills for reading?

Improvements in the effectiveness of reading instruction have not come about by variations in method per se. These variations have too often been based on guesses about the reading process. More definitive knowledge about the process and its component skills might lead to improvements that have to date eluded us. Tests—readiness, achievement, and diagnostic—should suggest directions for research, since they are designed to measure component skills.

Accordingly, it is the purpose of this paper to ask; what skills are required to perform well on current reading tests? An answer to this question calls for a critical evaluation of existing tests, many of which do not seem to examine reading ability by any definition of reading. Instead, both readiness and achievement tests appear to measure general language competence appropriate to middle-class caucasian families, and the effects of other kinds of preschool training.

Tests play an important role in beginning reading instruction, and necessarily so. Test performance determines choice of curriculum program for a child, the vocabulary to which he is exposed, and the attitudes and expectations of the teacher toward him.[1] Ideally, test performance should present information to the teacher about specific disabilities, information that can dictate the most efficient corrective action. The design and suggested use of readiness and achievement tests fits naturally into the analysis of beginning reading as an integration of component skills.

Unfortunately, there is little understanding of the reading process to which reference is frequently made. There are not adequate data on the stimulus cues used by readers at various levels of competence and stages of development. It is not known how these cues are selected and integrated during oral reading.[2] We don't really understand how basic skills (speaking, seeing and hearing) or more complex abilities (the child's linguistic fluency as measured by productive or recognition vocabulary, or the length and complexity of sentences) enter into the development of reading competence, however defined.

TESTS AND READING

Kindergarten children are a motley crew. They differ tremendously in height, weight, physical features, and intellectual capacity and potential. Some children will have already learned much of what is supposed to be taught in kindergarten and the first grade, while others will not have this advantage.[3] The ideal educational system meets each child at his level of competence and leads him as far as possible in the direction of the desired instructional goals. In this ideal system, tests serve an essential role in initial evaluation of a child and in continuing appraisal of the results of instruction. As Stott and Ball (1965) so nicely phrase it, "The assessment and equitable social management of individual differences in mental ability [are] matters of great practical importance [p. 4]." There is a special need to provide more effective assistance to children from culturally-disadvantaged backgrounds for whom present programs of testing and teaching seem especially inappropriate.

---

[1] Goslin (1968) points out some sociological problems related to standardized ability tests. "One of the most important criticisms of tests is that they contribute to their own validity by functioning as self-fulfilling prophecies. . . The likelihood of the optimistic prediction made on the basis of a high test score is . . . increased because the person who scores high receives special advantages, whereas the individual who does poorly is often denied opportunities."

[2] Goodman's (1968) work on oral reading errors represents an important step in this direction. Interesting possibilities are also implicit in the research on visual search (Neisser, 1968).

[3] Durkin (1966) studied the progress of early readers, children who already read at the first-grade level when they entered first grade. Although they generally had high IQ's, when matched on IQ they still maintained the one-grade advantage as late as the fifth and eighth grades. Durkin stated that attitude and the home environment were as important as instruction per se.

Readiness and achievement tests typically consist of a collection of subtests, each of which is equated, in name at least, with a unique subskill. For example, in the Metropolitan Readiness Test (Hildreth, Griffiths, & McGauvran, 1965), one finds the following list of subskills:

a. comprehension and use of oral language,
b. visual perception and discrimination,
c. auditory discrimination,
d. richness of verbal concepts,
e. general mental ability; capacity to infer and to reason,
f. knowledge of numerical and quantitive relationships,
g. sensory-motor abilities of the kind required in handwriting, writing of numerals and drawing,
h. adequate attentiveness; the ability to sit quietly, to listen and to follow directions.

Diagnostic reading tests usually stress that their purpose is not evaluation of overall reading performance, but determination of those specific skills in which a child has deficiencies. According to the Doren Diagnostic Reading Test (1956), "In an achievement test, the number of correct responses is the measure of the degree of success. In a diagnostic test, it is the mistakes which an individual makes that will indicate his areas of need, and an exact identification of the types of error will direct the examiner to specific remedial work [p. 17]." In the Durrell Analysis of Reading Difficulty (1955) a similar rationale is expressed. "Some of the common difficulties in learning are: (1) lack of adequate background abilities to perform the task, (2) failure to master the early elements on which later abilities are based, and (3) confusions resulting from instruction not correctly adjusted to the level of ability and the learning rate of the child, etc."

In fact, differences in the format of diagnostic, readiness, and achievement tests are minimal. All are comprised of three or more subtests, each designed to evaluate a different subskill assumed to be important in reading. The teacher is usually advised to consider not only the overall score in readiness or achievement tests, but to look at subtest performance for specific weaknesses. Given present teaching loads, such advice seems impractical.

Furthermore, closer examination reveals that the intercorrelations between subtests are so high that doubts are raised about whether independent skills are being tested, or (as an alternative hypothesis) whether the various skills related to reading develop at significantly different rates within the typical individual.

The Metropolitan Readiness Test (METRO) (Hildreth et al., 1965) consists of six subtests. Word Meaning, "a measure of the child's store of verbal concepts," is a picture vocabulary test with words chosen from kindergarten and primary word lists. Listening "strives to tap the child's ability to comprehend syllables and sentences." It is also a picture test. Matching requires the child to discriminate and perceive correspondences between word forms. Alphabet requires the child to recognize letters of the alphabet spoken by the examiner. Numbers tests familiarity with various arithmetic and number concepts. Copying is a test of the child's ability to copy letter-like forms.

In Table 1 are the subtest correlations from the standarization of the METRO.[4] The intercorrelations are all substantial. Factor analysis of the data in Table 1 indicated that two orthogonal components accounted for about 70% of the variance. Tests 1, 2, and 5 loaded on one factor; Tests 3, 4, 5, and 6 on the other. It is not obvious how one would interpret these factors, but they are not basic skills in any obvious sense.

TABLE 1

Intercorrelations Among Metropolitan Readiness Subtest Scores, N = 12,225

Subtest	2	3	4	5	6
1. Word Meaning	.83	.56	.59	.72	.49
2. Listening		.65	.60	.76	.52
3. Matching			.61	.71	.55
4. Alphabet				.74	.50
5. Numbers					.59
6. Copying					

Note.— Reproduced from Metropolitan Readiness Tests. Copyright 1965 by Harcourt, Brace & World, Inc. Reproduced by special permission of the publisher.

Another way of determining whether independent skills are being measured is to look at correlations between the METRO subtests and other criterion measures. During the standardization of the METRO, the Pintner–Cunningham Primary Test (P–C; Pintner, Bess, & Durost, 1946) and Murphy–Durrell Reading Readiness Analysis (M–D; Murphy & Durrell, 1964) were also

[4] The correlations in Tables 1 to 7 have been corrected for attenuation using reliability coefficients in the test manuals where possible.

administered. The intercorrelations are presented in Table 2. It can be seen that (a) the P–C is highly correlated with all METRO subtests, (b) the M–D Learning Rate subtest appears to measure something different from any of the METRO subtests, and (c) the other two M–D subtests correlate highly with the METRO. Except for the METRO Alphabet and M–D Letter Names subtests, which are identical, there is no evidence that subtests of the two readiness tests allow differential evaluation of basic abilities. For example, the correlation between the Listening and Phonemes subtests, both presumably sensitive to auditory discrimination, is about the same as Alphabet and Phonemes.

Data are also provided in the METRO manual on predictive validity with the Metropolitan Achievement and Stanford Achievement Tests (Table 3). None of the intercorrelations differ significantly from one another. The most reliable predictors of performance on any of the achievement subtests were Alphabet and Numbers subtests.

The Murphy–Durrell Reading Readiness Analysis (Murphy & Durrell, 1964) consists of three subtests. In the Letter Names test, the child must identify upper or lower case letters as the teacher gives their names. The Phonemes test is unique, in that the child is first taught to segment initial and final consonant sounds, and then is tested on segmentation ability. For example, the child hears the words, salt, sand, and soft as examples of the initial /s/ and then must mark those pictures whose names begin with /s/, e.g., sun, pillow, soap, and basket. The teacher reads the names of the pictured objects, so the effects of familiarity with the pictured objects should be negligible. In the Learning Rate test, the child is first taught to associate the names of common objects with their written equivalents. For example, the teacher writes on the board tongue, hair, and eyes and then names each word. After this preliminary session, the children are retested. The teacher pronounces one of the words and the child must pick out the spelled word from a list. Except for the Letter Names test, the M–D would appear to be tapping different abilities than the METRO and yet, as noted above, the various subtests of the METRO correlate highly with all but the Learning Rate subtest. In Table 4 are subtest correlations from the M–D standardization. In Table 5 are correlations with the Stanford Achievement Test. The data speak for themselves. The Learning Rate subtest has the lowest predictive validity, a strange result since this subtest involves procedures quite similar to those used in reading instruction.

Finally, consider the MacMillan Reading Readiness Test (Harris, 1963), which has four subtests. The Rating Scale consists of a subjective evaluation of the pupils readiness by the kindergarten teacher. Visual Perception requires matching of single letters or words. Auditory Perception measures ability to hear similarities and differences in initial consonant sounds and rhyming endings. This is also a matching test, based on pronunciations of key words by the teacher. Vocabulary and Concepts is a picture vocabulary test. In Table 6 are presented intercorrelations for two standardization groups, middle-class first graders and lower socioeconomic Negro children. Again, the intercorrelations are reasonably high for both populations. It might be noticed that the

TABLE 2

Correlations of Subtest Scores of Metropolitan Readiness with S[illegible]test Scores on Murphy–Durrell and Pintner–Cunningham Primary Tests, N = 12,225, inter-test interval 2–3 weeks

Test	Metropolitan Readiness Tests					
	Word Meaning	Listening	Matching	Alphabet	Numbers	Copying
Pintner–Cunningham Primary	.72	.82	.67	.60	.75	.61
Murphy–Durrell Analysis						
Phonemes	.60	.61	.54	.60	.64	.50
Letter Names	.58	.58	.57	.91	.70	.49
Learning Rate	.33	.35	.34	.37	.37	.30

Note.—Reproduced from Metropolitan Readiness Tests. Copyright 1965 by Harcourt, Brace & World, Inc. Reproduced by special permission of the publisher.

TABLE 3

Predictive Validity of Experimental Edition of Metropolitan Readiness Tests

Metropolitan Readiness Subtest	Metropolitan Achievement Test: Primary I[a]			
	Word Knowledge	Word Discrimination	Reading	Arithmetic Concepts & Skills
1. Word Meaning	.53	.45	.48	.52
2. Listening	.56	.50	.54	.57
3. Matching	.55	.50	.54	.52
4. Alphabet	.69	.66	.62	.52
5. Numbers	.65	.59	.63	.68
6. Copying	.45	.42	.45	.44

Stanford Achievement Test: Primary, Form I[b]

	Paragraph Meaning	Word Meaning	Spelling	Arithmetic Reasoning	Arithmetic Computation
Metropolitan Total Test	.58	.64	.74	.69	.64

Note.—Reproduced from Metropolitan Readiness Tests. Copyright 1965 by Harcourt, Brace & World, Inc. Reproduced by special permission of the publisher.

[a]Correlations based on medians from six groups of students, N per group ranging from 191 to 246.

[b]Correlations based on N = 96.

TABLE 4

Intercorrelations Among Murphy-Durrell Subtests, N = 12,231

Subtest	2	3
1. Phonemes	.62	.58
2. Letter Names		.40
3. Learning Rate		

Note.—Reproduced from Murphy-Durrell Reading Readiness Analysis. Copyright 1965 by Harcourt, Brace & World. Inc. Reproduced by special permission of the publisher.

rating by the kindergarten teacher is the best single predictor of test performance. One can do just about as well by asking the teacher to rate a child's readiness as by administering the entire test.

A comprehensive set of reading subtest intercorrelations is in the Bond and Dykstra (1967) report. All students were given the METRO, M–D, P–C, and Stanford Achievement tests. The Thurstone Pattern Copying Test and the Thurstone–Jeffrey Identical Forms Test were

TABLE 5

Predictive Validity Coefficients for Murphy-Durrell Reading Readiness Analysis with Stanford Achievement Test

Murphy-Durrell	Stanford Achievement: Primary I		
	Word Meaning	Paragraph Meaning	Word Study Skills
Phonemes	.67	.64	.70
Letter Names	.60	.61	.60
Learning Rate	.52	.54	.43

Note.—Reproduced from Murphy-Durrell Reading Readiness Analysis. Copyright 1965 by Harcourt, Brace & World, Inc. Reproduced by special permission of the publisher.

also administered to test copying ability and visual perception. In Table 7 are the subtest

TABLE 6

Subtest Intercorrelations for MacMillan Readiness Test

Subtest	Test I	Test II	Test III	Test IV	Total Score
I Rating Scale		.50	.43	.56	.96
II Visual Perception	.57		.43	.56	.76
III Auditory Perception	.67	.68		.48	.64
IV Vocabulary and Concepts	.54	.60	.69		.79
Total Score	.98	.76	.90	.69	

Note.— Upper set of r's, Disadvantaged Group, N = 142. Lower set of r's, Middle-class Group, N = 165.

TABLE 7

Intercorrelations Among Subtests Administered Before and After First-Grade Reading Instruction using Basal Programs, Bond and Dykstra (1967), N = 4,266

Subtest	2	3	4	5	6	7	8	9	10	11	12	13
1. M–D Phonemes	.52	.42	.35	.29	.43	.38	.50	.54	.50	.51	.40	.53
2. M–D Letter Names		.49	.31	.30	.41	.30	.46	.60	.56	.46	.51	.55
3. M–D Learning Rate			.28	.27	.33	.37	.38	.44	.45	.34	.35	.40
4. Thurstone Copying				.32	.26	.25	.49	.34	.34	.32	.30	.36
5. Thurstone-Jeffrey Identical Forms					.28	.24	.46	.29	.29	.32	.26	.31
6. METRO Word Meaning						.77	.51	.42	.38	.61	.33	.41
7. METRO Listening					.51	.33	.34	.49	.24	.38		
8. P–C Raw Score						.50	.47	.59	.35	.49		
STANFORD ACHIEVEMENT:												
9. Word Reading							.87	.62	.71	.81		
10. Paragraph Meaning								.58	.73	.77		
11. Vocabulary									.47	.67		
12. Spelling										.71		
13. Word Study Skills												

Note.— This table is based on an N of 4,266 from 187 classes in 17 projects. Reliability coefficients for Tests 4 and 5 were not available, and hence correlations associated with those tests are not corrected for attenuation.

intercorrelations.[5] Subtest correlations between and within tests were relatively high, except for the METRO Listening and Thurstone-Jeffrey Identical Forms Tests, which for these students were also unrelated to the criterion performance on the Stanford. The P–C vocabulary test correlated to the same extent with all of the readiness and achievement subtests. Factor analysis showed that two factors accounted for 55% of the variance in Table 7. The first factor loaded most heavily on Tests 2, 9, 10, 12, and 13. It appears that (a) ability to identify the letters of the alphabet and reading achievement at the end of first grade are closely related, and (b) four of the five subtests on the Stanford yield similar achievement scores. The

[5] The data in Table 7 are based on students taught by some form of basal reading program. Correlation matrices for students taught by different reading programs, such as ITA, language experience, linguistic/phonics, etc., were quite similar. Table 7 is representative, and the original report may be consulted for details. Dykstra (1967) has reported that data from the same children at the end of the second grade yield a similar pattern or results.

second factor loaded on Tests 6, 7, and 8, which are all vocabulary tests. Interestingly, knowledge of the letters of the alphabet at the beginning of first grade predicted reading achievement at the end of first grade as well or better than vocabulary at the beginning of first grade, even though these children were taught and tested by procedures which would stress comprehension.

We have not chosen these particular readiness tests with any malicious intent. To the contrary, they appear to constitute the most adequately constructed and standardized readiness tests available. One might conclude that it is difficult to identify separable skills since, on the face of it, different testing procedures and materials are represented in the collection of subtests. An alternative interpretation is that perceptual and cognitive development are such that an individual child is not likely to differ much in the degree to which he has mastered the requisite perceptual and language skills.

LANGUAGE AND IQ TESTING

Psychologists have for some time faced problems analogous to the measurement of independent reading skills in the assessment of intelligence and language ability. The first IQ test was developed by Sir Frances Galton to test his theory of inherited intellect. Galton devised a battery of tests measuring sensory-motor performance, immediate memory, and other primary skills, but was discouraged to find that none of these measures bore any substantial relationship to other criteria of intelligence. Alfred Binet in France was more successful in devising tests with immediate practical implications; they predicted school performance. Binet's approach proved viable in applied settings, whereas the research tradition begun by Galton finds its current niche in the experimental psychology laboratory.

Intelligence tests, like reading readiness and achievement tests, generally consist of subtests designed to test presumably independent cognitive functions. The question of the relative independence of the cognitive abilities tapped by the various subtests has been important both practically and theoretically. For example, the Wechsler Intelligence Scale for Children (1949) consists of two scales, Verbal and Performance. The Verbal scale is designed to measure language fluency, and the Performance scale, sensory-motor and perceptual ability. The subtests for the two scales are quite different. A typical Verbal item is, "What is the population of the United States?", while a typical Performance subtest requires a child to put together a simple jigsaw puzzle. Correlation between the two scales is about .80 (corrected for attenuation, about .86). The high correlation is useful for diagnosis; large differences between Verbal and Performance scores are presumed indicative of abnormal intellectual functioning. In the typical child, however, the Verbal and Performance subtests produce very nearly the same score.

Construction of tests sensitive to identifiable skills has been important for investigation of the "differentiation hypothesis" (cf. Stott & Ball, 1965). The supposition is that early in development, prior to age three, cognitive abilities are not differentiated to any extent. As a child matures, cognitive abilities may develop at unequal rates and hence appear as differentiated. The usual test of the hypothesis has relied on correlations among subtests designed to measure different skills. A report by Meyers, Dinzman, Orpet, Sitkei, & Watts (1964) is representative. These investigators constructed tests for children between two and six years of age which were designed to measure four types of basic cognitive abilities: psychomotor, perceptual speed, linguistic ability, and figural reasoning. They were interested in two questions. First, did the subtests measure independent identifiable abilities, or could the data be adequately described by a general intelligence factor? Second, did the degree of skill differentiation increase with age? Meyers et al, were reasonably successful in constructing subtests sensitive to separate cognitive skills. While the intercorrelations were not as low as one might desire (range .04 to .57, median .34), factor analysis showed that the data were adequately described by four factors. There was no support for the differentiation hypothesis.

A study by Lesser, Fifer, and Clark (1965) provides another example. These investigators sought to determine whether or not children from different social classes and cultural groups in New York City exhibited unique patterns of mental abilities. They constructed subtests to measure Verbal, Reasoning, Number, and Space ability. Moderate subtest intercorrelations within social and ethnic groups were observed (range .12 to .72, median .35). The Reasoning subtest correlated most highly with the other subtests, especially Number and Space. Lower-class children performed more poorly than middle-class children on all subtests in all ethnic groups (percentile difference of about 10 points on the average). Chinese and Jewish children performed better on the Reasoning, Number, and Space scales than Negroes and Puerto Ricans, but performance on these subtests was not substantially different within

those subgroups. The Jewish children showed better and the Puerto Ricans poorer Verbal ability than did the Chinese and Negro children, whose scores were similar. Performance on the Verbal scale was thus different from performance on the other three, but there was less convincing evidence that the nonverbal subtests were measuring substantially different abilities.

As a final example, consider the Illinois Test of Psycholinguistic Ability (ITPA) (McCarthy & Kirk, 1963) which is based upon a rather elaborate model of language functioning. As children mature, presumably new language skills are added and old skills become further refined. The original test was composed of nine subtests, selected to measure skills at several points within the language system. It was standardized on children ranging in age from 2.5 to 9 years. Subtest intercorrelations were generally high for all age groups, and factor analysis showed that most of the systematic variance in the tables of intercorrelations could be accounted for by a single variable, best described as general linguistic ability. There was no consistent evidence of a systematic development of specific skills with age.

Further work on the same test by Quereshi (1967), using a different factor analytic technique, led to a more optimistic outcome. The relative importance of the general factor appeared to decrease with age (41% of the variance at age 2.5 to 23% at age 9), and three group factors were found, each accounting for 10–15% of the variance. The analysis was based on a rational division of the subtests on the ITPA into subsets, and hence made more sense than the unconstrained analysis of McCarthy and Kirk. The procedure did not yield orthogonal factors, and correlations among the four factors (the general factor and the three group factors) were about .45. Quereshi concluded that, because of the importance of the general factor and the high factor correlations, test constructors should concentrate on the general factor. However, again there is evidence that tests can be constructed that measure component skills that are to some degree separable.

THE TROUBLE WITH TESTS

Tests can be used for several purposes—prediction, diagnosis, measurement of aptitude, interest, performance or achievement. In the area of beginning reading, there is little trouble in finding tests to predict performance or reading achievement at the end of first grade. As mentioned previously, a child's ability to name the letters of the alphabet and the kindergarten teacher's ratings are both reliable predictors. Correlation continues to resist any efforts to be equated with causality, however. By the end of first grade, most children have learned to identify the letters of the alphabet, but many have not become satisfactory readers (e.g., Olson, 1958). Children who are not able to handle phonetic discrimination or segmentation are also likely to be poor readers (Durrell & Murphy, 1953). The conclusion has been drawn that such children must be taught to listen more carefully to what they hear and say. Yet pilot studies in our laboratory and the experience of teachers with whom we have spoken suggest that it is difficult to explain phonetic segmentation to a child until he learns to read. As soon as the child learns the reading game, i.e. the correspondence between letters and sounds, he acquires a vocabulary which allows him to talk about phonetic segmentation. This is not meant to imply that a nonreader cannot be taught to segment. The question is whether segmentation ability is a prerequisite to reading, or vice versa.

Performance on readiness subtests has been put forward as a source of diagnostic information, yet there is no clear evidence of the validity of these measures as diagnostic indicators, nor is it apparent what remedial action should be taken when a child performs poorly on a readiness subtest. The trouble with reading readiness tests is that they do not provide measures of component skills that are related to reading performance in any well defined manner.

There has been relatively little effort to establish the validity of diagnostic test procedures in remedial reading. The causal relation between a particular deficiency and reading is established either by fiat or through correlational evidence. For example, it is considered obvious that if a child cannot articulate correctly, he will therefore have problems in learning to read. Accordingly, speech therapy is recommended. To the best of our knowledge, there is little evidence of high correlation between articulation and reading achievement, nor has it been shown that correction of articulation per se has any positive effect on reading performance.

TEST-TAKING AND LANGUAGE SKILLS

The inclusion of different types of subtests in readiness tests would seem defensible to the extent that the subtests are sensitive to different skills and insensitive to general abilities. Yet there is reason to suspect that current tests are so constructed that two general ability factors determine whether a child can perform well on any subtest. The first of these

factors is the ability of the child to follow instructions, and the second (and related) factor is general language competence. These characteristics of the test may be appropriate and useful in prediction. The problem is that they compromise the diagnostic value of the test.

Consider some specific examples from the Metropolitan Readiness Test. The first subtest, Word Meaning, is a picture vocabulary test in which the pupil selects from three pictures the one corresponding to a word spoken by the teacher. Presumably, the subtest is designed to measure extent of recognition vocabulary. The words were selected from standard kindergarten and primary word lists. Yet from the construction of the subtest, the selection of target items and alternatives, it is hard to ascribe performance to extent of recognition vocabulary alone. Of the sixteen target items, eight (windmill, moose, yarn, knitting, toboggan, spectacles (not glasses), blueberry, and moccasin) are either archaic, specialized, or unfamiliar.[6] What remedy is prescribed when a child does poorly on this subtest? The selection of alternatives is likewise curious (the target item is underlined): walnut, chestnut, acorn; shingled house, brick house, stone house; knitting or tatting a bootee, knitting (a larger item), embroidery; hoop, horseshoe, hoof. The ability to select the correct item depends on visual discrimination and logical inference as much as vocabulary. The choice of vocabulary items appears singularly inappropriate for urban children, especially those from lower socioeconomic backgrounds.

Similar comments hold for the Listening subtest, which is largely a test of inferential ability, attention to visual detail and memory. For example, "in the fall, father rakes the leaves and burns them;" the child is to distinguish between a man lighting a fire in a brick barbecue, a man tossing leaves into a basketful of burning leaves, and a man raking leaves onto a burning pile. Or again, "It is a big animal. It has four legs like other animals. It has a tail. It has many things other animals have, but it has one thing they do not have." Pictured are a bear, a horse and an elephant. The test is designed to measure ability to comprehend sentences. If a child performs poorly on this test, what should be done?

---

[6] Except for windmill and knitting, these words are relatively rare. According to the Thorndike-Lorge (1944) count, they are not among the 5,000 most common words in English. The same comment holds for the items from the Metropolitan Achievement Test mentioned later in the paper, where only bonnet is among the 5,000 most common words.

Investigation of other readiness tests turns up similar examples. In the Lee-Clark Reading Readiness Test (1962), the kindergartener is asked to identify a "short-haired dog"; a Doberman, a Saint Bernard, and a cocker spaniel are pictured, but small and with little detail. For another item, the instructions are, "Put a mark on the two little chickens." The alternatives are a hen and a pair of chicks, two medium-sized chickens with combs and wattles, and a slightly larger pair (a hen and a rooster, judging from the tail feathers on one). The two middle-sized birds are the correct choice. For another item, the child must indicate which vehicle carries the most people — a horse, a jet airplane, a car, or a boat.

The Lee-Clark predicts first-grade reading achievement reasonably well. Data presented in the manual show that at the end of first grade, those children who did most poorly on the readiness test can be expected to be half a grade behind their classmates. One can further predict that these children will be at a relatively greater disadvantage in later grades. The sad fact seems to be that readiness test information can be used only to delay the beginning of reading instruction by intervention of "readiness" activities.

Achievement tests, frequently used as criterion devices, are also inadequate. Bormuth (1968) has argued that "achievement tests constructed by current methods have no logical and objectively demonstrable relation to the instruction . . . a score on an achievement test made by (current) procedures must be interpreted as a student's response to the test writer's responses to the instruction."

There are three reading subtests on the Metropolitan Primary Achievement Test. The Word Knowledge test is designed to measure the child's sight vocabulary or word recognition ability. It is a picture vocabulary test. In the Word Discrimination test, the teacher pronounces a word and the child must then mark the corresponding word from a list of four. In order to perform well on this test, the child must be able either (a) to associate the pronounced word with its printed representation, and choose from a set of words that are visually similar, or (b) remember the word while pronouncing each of the test items and comparing with the test item. The third test, Reading, requires the child to look at a picture, decide what the picture portrays and then select the sentence which best describes what is happening. (One of the items requires the child to infer that a man in a blue suit who helps children and tells them to stop and go, is a policeman.)

All of these subtests require of the child a fair degree of inferential ability, an extensive

reading vocabulary (e.g., muss, mane, wringer, bonnet, clothespin), and the ability to discriminate very sharply between conceptually similar and conceptually ambiguous items (e.g., a picture of a turtle going down a road past a sleeping rabbit— "the turtle is afraid the rabbit will get ahead of him" or "the rabbit sleeps while the turtle crawls down the road"). There is no question that a bright child who reads well can perform well on all of these tests, or that a dull child who can't read will do poorly. On the other hand, none of the tests constitute the most straightforward test of the child's reading ability, whether one chooses to stress the decoding or comprehension aspects of reading. The Word Discrimination test is as much a test of spelling ability and the clarity of the teacher's articulation as it is the child's ability to read. A child might be able to read aloud every word in the test and still perform very poorly. To be sure, this is an achievement test, not a diagnostic test. The question remains, is this the best approach to the design of an achievement test, and must the design of achievement tests be such that they provide no useful diagnostic information?

MEASUREMENT OF COMPONENT SKILLS

At the beginning of this paper, the question of research priorities was raised. In answer, it has been suggested that substantial improvements in reading instruction will require more detailed knowledge of the reading process and the component skills which relate to the development of this ability. There is an obvious need for more adequate measures of basic skills.

For the past two years in our research program at the Wisconsin Research and Development Center, we have been investigating articulation and phoneme discrimination skills in young children— preschoolers, kindergarteners and first graders. We quickly discovered that a major hurdle was development of testing procedures that made minimal demands on the child, apart from the skill being measured.

The usual approach in constructing an articulation test (e.g., Templin, 1957) has been to select pictures of familiar objects until all the major phonetic contrasts in English are included in the set. A child is shown each picture and asked to name the object. To do well on the test, a child must (a) be able to interpret an abstract representation of an object, (b) be familiar with the object in question (i.e., recognize it and have an appropriate name for it in the speaking vocabulary), and (c) be able to give the pronunciation correctly. Since the objects are presumed to be familiar, this type of testing procedure is sensitive to dialect variations. Problems of recognition, familiarity, and dialect should be minimized if the child repeats a word spoken by the experimenter or recorded on tape. In fact, Snow and Milisen (1954) showed that articulation performance of children with speech defects was significantly better on imitation than on picture-naming tests using the same words.

We have just completed testing the articulation ability of over 600 kindergarten and first-grade children using an imitation procedure (Venezky and Calfee, 1968). The details are beyond the scope of this paper, but certain findings are germane to the discussion.

In the testing, a child repeated each word as it was read by a trained experimenter or played from a tape recorder. As might be expected, the recorded presentation produced a somewhat higher error rate, but the variability between schools was twice as great for the live presentation. It appears that even a trained tester may contribute to a child's articulation performance.

There were marked word-context effects even though an imitation procedure was employed. For example, there were 116 errors on the /br/ cluster in broil, but only 21 in breathe. Initial /b/ was mispronounced 3 times in birth, but 40 times in beige. These differences might be attributed to familiarity or word frequency, but not the /k/ errors in coins (3) vs. cage (32). Context effects are especially noteworthy in light of the finding that children did not produce uniform patterns of errors. Most of the errors involved semivowels, /r/ and /l/, or fricatives, /s/, /z/, /θ/ and /ð/. A child might make a substitution such as /w/ for /r/ in one context but not in others. In less than 5% of the children was there evidence that a child was totally unable to produce one or more phonemes.

Substitution or deletion of initial /s/ is fairly common in first graders, especially when the sibilant is part of a cluster. To get a clearer picture of the consistency of /s/ errors, a 30-item test was prepared consisting of consonant clusters such as /sp/, /sk/, /sw/, etc. Typical items were span, speak, spright, spray, sprawl and spring. Of 57 children, 18 made at least one /s/ error; of those 19, only 2 made more than 4 errors, and none missed all of the 30 items. Thus, even with a difficult phoneme, many children make occasional errors, but very few children are entirely incapable of producing the phoneme.

Our data suggest that the phonetic environment may be as important a determinant of performance as individual differences. For example, in clusters such as /br/, /pr/, and /fr/ where a front (bilabial or labio-dental)

GPO 909-594-3

consonant was followed by a central semivowel, the semivowel was replaced more frequently (usually by /w/) than where the consonant was central as in /tr/ and /dr/. The transition from front to central, which covered a relatively long articulatory distance, was difficult for these children. That /r/ was correctly pronounced in a less difficult context (from a motoric standpoint) suggests that a child may be able to discriminate between /r/ and /w/ quite well, even though occasional replacements occur.

This result is noteworthy because of the frequent assertion that speech problems reflect (phonetic) discrimination failures. Templin's (1957) finding of substantial correlations (.4 to .7) between articulation and discrimination has been taken as support for the assertion. Once again, correlation is not causality. The relation which remains to be established is the existence of articulation and discrimination problems common to specific phonemes or articulatory features.

Phonetic discrimination has been shown to be related to reading performance. The usual testing procedure with children has been a same-different task (e.g., Robbins & Robbins, 1948). The child is presented with a pair such as /fa/-/θa/ and asked whether the two items were the same. Unfortunately, the concept of identity is not well developed in all kindergarten and first grade children. Although many children use the words same and different or alike and unalike, their interpretation of these terms when applied to speech may be different from the experimenter's interpretation. We ran headlong into this problem early in our testing program when one of the children replied "different" when shown two cards containing identical geometrical forms. When asked to justify his answer, the child pointed out that one of the cards had a smudge on it. With older or more test-sophisticated children, it is easier to communicate the dimensions with regard to which identity is to be judged. With younger children, or where the material being tested may pose a new and difficult test for the child under the best of circumstances, the relevant dimensions may be extremely difficult to interpret for the child.

In another type of phonetic discrimination test, the child is asked to determine whether or not a criterion phoneme such as /s/ is present in a familiar word. For example, the child hears exemplars such as sun and soup and then is required to point to those pictures in a list which contain the same initial sound. The child must be able to recognize pictures, identify the objects in them, segment and abstract the relevant phoneme, and discriminate between phonemes. There is no reliable information about how a child who substitutes /θ/ for /s/ performs when he is asked to mark words beginning with /s/. In any event, errors on this type of task may be traced to many sources.

CONCLUSION

Like others, we would like to find more effective ways to teach reading. It seems futile to introduce more new methods until necessary insights into the nature of the reading process are established by appropriate research. Dissection of the process into its components is an impossible venture when each measure correlates with every other measure to the same extent; hence our concern with testing procedures. We are optimistic about the possibility of finding reliable instruments sensitive to well defined skills. We would be quite pleased to find tests of articulation and discrimination ability were only slightly correlated with one another, and that neither was significantly related to performance on current readiness or achievement tests. After all, it is hard to believe that the sum total of a child's intellectual ability can be measured by his knowledge of the letters of the alphabet prior to first grade. Reading is a vital skill without which a child cannot succeed in virtually any other area. Today, it is possible to predict quite reliably those children who are not going to make it. This damning prediction must be changed into a prescription for treatment.

REFERENCES

Bond, G. L., & Dykstra, R. The cooperative research program in first-grade reading instruction. Reading Res. Quart., 1967, 2, Num. 4 (Complete).

Bormuth, J. On the theory of achievement test items. Mimeographed paper, Univ. of Chicago, 1968.

Chall, G. Learning to read: the great debate. New York: McGraw-Hill, 1967.

Doren, M. Doren Diagnostic Reading Test. Minneapolis: Educational Test Bureau, 1956.

Durkin, D. Children who read early. New York: Teacher's College Press, Columbia Univ., 1966.

Durrell, D. D. Durrell analysis of reading difficulty. New York: Harcourt, Brace, & World, 1955.

Durrell, D. D. & Murphy, H. A. The auditory discrimination factor in reading readiness and reading disability. Education, 1953, 73, 556–560.

Dykstra, R. Continuation of Coordinating Center for first-grade reading instruction programs. (Final Report of Project 6-1651, Contract OEC3-7-001651-04R), Minneapolis: Univ. of Minnesota, 1967.

Goodman, K. S. Reading: a psycholinguistic guessing game. J. Reading Specialist, 1968, in press.

Goslin, D. A. Standardized ability tests and testing. Science, 1968, 159, 851–855.

Harris, A. J. MacMillan reading readiness test. New York: MacMillan, 1963.

Hildreth, G., Griffiths, N. & McGauvran, M. Metropolitan readiness tests. New York: Harcourt, Brace & World, 1965.

Lee, J. M., & Clark, W. W. Lee-Clark reading readiness test, manual. New York: McGraw-Hill, 1962.

Lesser, G. S., Fifer, G., & Clark, D. H. Mental abilities of children from different social-class and cultural groups. Monog. Soc. Res. Child Dev., 1965, 30, Whole No. 4.

Levin, Harry. "Reading research: What, why, and for whom?" Elementary English, 1966, 43, 138–147.

McCarthy, J. J. & Kirk, S. A. The construction, standardization and statistical characteristics of the Illinois test of psycholinguistic abilities. Madison, Wisc.: Photo Press, Inc., 1963.

Meyers, C. E., Dinzman, H. F., Orpet, R. E., Sitkei, E. G., & Watts, C. A. Four ability-factor hypotheses at three preliterate levels in normal and retarded children. Monog. Soc. Res. Child Dev., 1964, 29, Whole No. 4.

Murphy, Helen, & Durrell, D. D. Murphy-Durrell reading readiness analysis. New York: Harcourt, Brace & World, 1964.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1968.

Olson, A. V. Growth in word perception abilities as it relates to success in beginning reading. J. of Educ., 1958, 140, 25–36.

Pintner, R., Bess, C., & Durost, W. Pintner-Cunningham primary test, form A. (general ability tests, revised). New York: Harcourt, Brace & World, 1946.

Quereshi, M. Y. Patterns of psycholinguistic development during early and middle childhood. Educ. psychol. Meas., 1967, 27, 353–365.

Robbins, S. D., & Robbins, R. S. Robbins speech sound discrimination test. Magnolia, Mass.: Expression Co., 1948.

Russell, D. H., & Fea, H. Research on teaching reading. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand-McNally, 1965, pp. 865–928.

Snow, K., & Milisen R. The influence of oral pictoral representation upon articulation testing results. J. Speech Hearing Dis., Monog. Suppl. 4, 1954, 30–36.

Stott, L. H., & Ball, R. S. Infant and preschool mental tests: review and evaluation. Monog. Soc. Res. Child. Dev., 1965, 30, Whole No. 3.

Templin, Mildred C. Certain language skills in children. Minneapolis: Univ. of Minnesota Press, 1957.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York: Columbia Univ. Press, 1944.

Venezky, R. L., & Calfee, R. C. The patterning of articulation errors in kindergarten and first grade. Paper presented at Psycholinguistics and Reading Presession, Intern. Reading Assn., Boston, April, 1968.

Wechsler, David. Wechsler intelligence scale for children, manual. New York: Psychological Corporation, 1949.

GPO 809-594-2