DOCUMENT RESUME

ED 194 575 TM 800 704

AUTHOR	Johnson, Dale D.; And Others
TITLE	Interim Report: The Refinement of the Test Battery to Assess Word Identification Skills. Report from the Project on Studies in Language: Reading and Communication.
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Individualized Schooling.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	WRDCIS-TR-544
PUB DATE	Jul 80
GRANT	OB-NIE-G-80-0117
NOTE	88p.
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Consonants; *Diagnostic Tests; Elementary Education; *Item Analysis; *Phonics; *Structural Analysis (Linguistics); Test Construction; Vowels; *Word Recognition
IDENTIFIERS	*Word Identification Test (Johnson et al)

ABSTRACT

The research discussed in this report is a continuation of an effort to develop a set of diagnostic subtests in phonics and structure with empirically determined mastery levels. The complete test will assess the word identification skills of elementary school children. Revisions made in the test battery following administration of the Phonics and Structure Subtests are documented. A study in which the revised subtests were administered to 80 grade 2 and 3 pupils is also described. Item analysis information for use in the preparation of the subtests for the final administration of the Word Identification Test battery is presented. These revisions will improve the diagnostic precision of the tests and eliminate concerns regarding correct response position and frequency of occurrence of specific distractor categories. Results of evaluating the grade level appropriateness of each subtest indicated: (1) the Phonics Subtests could be administered to younger children (grade 1); and (2) the Structure Subtests could be administered to older children (grades 4 and 5 for affixes, and grade 4 for possessives and contractions). (Author/RL)

Technical Report No. 544

INTERIM REPORT: THE REFINEMENT OF THE TEST BATTERY TO ASSESS WORD IDENTIFICATION SKILLS

by

Dale D. Johnson, Susan D. Pittelmen
Judy Schwenker and Linda F. Shriberg

Report from the Project on
Studies in Language: Reading and Communication

Dale D. Johnson
Faculty Associate

Wisconsin Research and Development Center
for Individualized Schooling
The University of Wisconsin
Madison, Wisconsin

July 1980

Published by the Wisconsin Research and Development Center for Individualized Schooling. The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Center Grant No. OB-NIE-G-80-0117

MISSION STATEMENT

The mission of the Wisconsin Research and Development Center is to improve the quality of education by addressing the full range of issues and problems related to individualized schooling. Teaching, learning, and the problems of individualization are given concurrent attention in the Center's efforts to discover processes and develop strategies and materials for use in the schools. The Center pursues its mission by

- conducting and synthesizing research to clarify the processes of school-age children's learning and development
- conducting and synthesizing research to clarify effective approaches to teaching students basic skills and concepts
- developing and demonstrating improved instructional strategies, processes, and materials for students, teachers, and school administrators
- providing assistance to educators which helps transfer the outcomes of research and development to improved practice in local schools and teacher education institutions

The Wisconsin Research and Development Center is supported with funds from the National Institute of Education and the University of Wisconsin.

WISCONSIN RESEARCH AND DEVELOPMENT
CENTER FOR INDIVIDUALIZED SCHOOLING

ACKNOWLEDGMENTS

The authors gratefully acknowledge the cooperation of the pupils and teachers at Belleville Elementary School in Belleville, Wisconsin. Their support of and participation in this research is greatly appreciated.

The authors also acknowledge with gratitude Lawrence D. Shriberg, Professor of Speech Pathology, University of Wisconsin-Madison; Dominic W. Massaro, Professor of Psychology, University of Wisconsin-Madison; Peter A. Schreiber, Professor of English and Linguistics, University of Wisconsin-Madison; and Richard L. Venezky, Unidel Professor of Educational Foundations, University of Delaware, for their helpful suggestions.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

ABSTRACT

The research discussed in this report is a continuation of an effort to develop a set of diagnostic subtests in phonics and structure with empirically determined mastery levels which will assess the word identification skills of elementary school children.

In spring 1977, a prototype of the test battery was developed and pilot tested on 282 pupils. Between winter 1977 and spring 1979, the subtests were administered to approximately 1700 elementary school students from five regions of the country. This report documents the revisions made in the test battery following a winter 78/79 and spring 1979 administration of the Phonics and Structure Subtests. The report also describes a study in which the revised subtests were administered to 80 second and third grade pupils.

The primary purpose of this study was to obtain item analysis information for use in the preparation of the subtests for the final administration of the Word Identification Test battery, which was conducted in spring 1980. While only minor revisions were necessary, these revisions should improve the diagnostic precision of the tests and eliminate concerns regarding correct response position and frequency of occurrence of specific distractor categories.

Another purpose of the present study was to evaluate the grade level appropriateness of each subtest. Results indicated (a) the Phonics Subtests could be administered to younger children (first grade), and (b) the

Structure Subtests could be administered to older children (fourth and fifth grade for Affixes, and fourth grade for Possessives & Contractions).

This report is the fourth report in a series of five. The final report in the series, which will be available in fall 1980, will mark the conclusion of the research conducted over the past three and a half years by the Project on the Assessment and Analysis on Word Identification Skills in Reading.

The work reported in this paper is a continuation of a research effort begun in spring 1977 by the Project on Assessment and Analysis of Word Identification Skills in Reading. The focus of the research over the last three years has been (a) to identify those word identification skills which appear to relate most closely to the comprehension of written language; and (b) to empirically determine mastery levels for each word identification subtest based on performance of groups of subjects stratified by comprehension ability.

The ultimate goal of the research is the development of a set of diagnostic subtests with empirically determined mastery levels, which will assess the word identification skills of elementary school children. The test battery consists of five subtests within two major skill areas: phonics and structural analysis. The Word Identification Test battery, with mastery levels set by comprehension standards rather than by an arbitrary cutoff score (such as 80%), is intended to assist teachers in making sound instructional decisions based on the accurate assessment of word identification skills.

In spring 1977, a prototype of the test battery was developed which assessed skills in three broad areas of word identification: phonics, structure, and context. In order to provide baseline information, the battery also included a section assessing reading readiness skills. The prototypes were pilot tested on a total of 282 pupils in grades one, three, and five. Following the data analysis, revisions were made on the individual subtests (Johnson, Pittelman, Schwenker, Shriberg, & Morgan-

Janty, 1978). In winter 1977 and spring 1978, the revised Word Identification Test battery (assessing skills in the areas of phonics, structure and contextual analysis) and the reading subtest of the Metropolitan Achievement Tests (Farr, Prescott, Balow, & Hogan, 1978) were administered to approximately 1150 second, fourth, and sixth grade public elementary school children from five regions of the United States (Johnson et al., 1978). Following analysis of the data from these test administrations, the criteria which guided test construction were evaluated, and additional criteria were incorporated in the development of the present tests (Johnson, Pittelman, Schwenker, & Shriberg, 1979; Johnson, Shriberg, Pittelman, & Schwenker, 1979). Furthermore, a decision was made to limit the development of the Word Identification Test battery to the areas of phonics (with Consonants and Vowels Subtests) and structural analysis (with Inflected Endings and Affixes Subtests).

In winter 78/79, the revised Inflected Endings Subtest was administered to 143 third and fourth graders, and the revised Consonants Subtest was administered to 187 third through fifth graders. In spring 1979, the revised Affixes Subtest was administered to 257 fourth and fifth graders and the revised Vowels Subtest was administered to 190 third through fifth graders (Johnson, Pittelman, Schwenker, & Shriberg, 1979; Johnson, Shriberg, Pittelman, & Schwenker, 1979).

This report documents the revisions made in the Phonics and Structural Analysis Subtests following the winter 78/79 and spring 1979 administration of the Phonics and Structure Subtests, and also describes a study in which

the revised subtests were administered to 80 second and third grade pupils. The primary purposes of the study conducted in fall and winter 79/80 were to obtain item analysis information prior to preparation of the final version of the tests and to evaluate the test directions and administrator's manual for each subtest.

In spring 1980, the final version of the Word Identification Test battery was administered to approximately 600 first through fifth grade elementary school students. The performance data will be used to examine correlations between reading subskills, as measured by the various subtests in the Word Identification Test battery, and global comprehension ability, as measured by a standardized reading test. In addition, the data collected in spring 1980 will be used to establish empirically based levels of skills mastery for each of the reading subskills assessed in the battery. Results of the above analysis will form the basis of the final Project monograph which will be published in fall 1980.

DEVELOPMENT OF THE STRUCTURAL ANALYSIS SUBTESTS

The structural analysis component of the Word Identification Test battery assesses students' skills in the areas of inflected endings, contractions and possessives, and affixes.

Method

SUBJECTS

A total of 80 second and third grade pupils from Belleville Elementary School participated in the study to evaluate the three structure subtests (see Table 1). Belleville, Wisconsin is a rural community approximately 30 miles from Madison. All four classrooms participating in the study were heterogeneously grouped.

STIMULI

The structural analysis component of the Word Identification Test battery is comprised of three subtests: revised versions of the Inflected Endings and Affixes Subtests and the newly created Contractions & Possessives Subtest.

Inflected Endings Subtest

A 39-item test of inflected endings was developed. Seven different target inflected endings were assessed by the revised subtest compared to eight assessed in the winter 78/79 version. The target inflected endings

Table 1

Structural Analysis Subjects by Grade

($\underline{N}$ = 80)

	Grade 2	Grade 3
Classroom A	21	
Classroom B	22	
Classroom C		19
Classroom D		18
Total per grade	43	37

assessed in each version, as well as the number of items for each target inflected ending, are presented in Table 2.

Analysis of the data from the winter 78/79 administration of the Inflected Endings Subtest to fourth and fifth grade students had shown a marked ceiling effect. Performance on items involving the possessive use of the apostrophe, however, showed significantly lower scores. Because of this inconsistency, a decision was made to delete 's and s' words both as items and as incorrect response choices from the Inflected Endings Subtest and to develop a separate subtest which would assess possessives and contractions as alternative uses of the apostrophe.

The elimination of words containing the 's and s' led to a second change in the Inflected Endings Subtest. Because 's and s' endings had been frequently used as incorrect response choices in the subtest, and because many of the root words in the test had too few inflected forms to generate replacement incorrect response choices, new root words had to be selected for 30% of the items.

Another consideration, the length of the item stems, led to a third type of revision in the Inflected Endings Subtest. These revisions were aimed at easing the reading load of the Inflected Endings Subtest by shortening sentences whenever possible while still retaining adequate context for the correct response. Sentence shortening would serve to make the subtest more appropriate for second and third grade students, the intended target population.

The format for the revised Inflected Endings Subtest remained similar to the format of the previous version of the test: Pupils saw a sentence

Table 2

Target Inflected Endings Assessed in Winter 78/79 and Winter 79/80 Versions of the Inflected Endings Subtest

	Number of Items	
Target Inflected Ending	Winter 78/79	Winter 79/80
s (plural)	6	5
(e)s (verb)	5	6
(i)ed	5	6
ing	5	3
's or s'	4	-
(i)es	3	-
er	3	3
est	3	3
y	-	1
Other Target Words		
tense marker (with vowel change)	-	1
root (correct response)	9	11
Total Number of Items	43	39

with a word missing and were instructed to select the response choice (provided below the sentence) that best completed the sentence. However, a fourth response choice, none of these, was added to the revised test to increase its reliability. The revised Inflected Endings Subtest, therefore, contains four response choices for each item. Response choices consisted of root words, root words plus inflected endings, and the response choice none of these. Figure 1 is a copy of the directions and practice items from the revised Inflected Endings Subtest.

Contractions & Possessives Subtest

A 31-item test was developed for the current study to assess children's knowledge of contractions and possessives. (In the previous test battery, contractions were not assessed and possessives items were incorporated in the Inflected Endings Subtest.) The rationale for assessing both contractions and possessives in the same test is that both constructions use the apostrophe, although to convey different meanings. The ability to distinguish between these two uses of the apostrophe is important for obtaining the intended meaning of connected text which, in turn, affects comprehension. Consider, for example, the following two sentences:

George's lead part in the school play made Tom jealous.
(Possessive form)

George's lead actor and Tom's assistant director in the school play. (contracted form)

To comprehend, a reader must be able to distinguish between the different meanings of the apostrophe in the word George's in these two seemingly similar sentences.

Inflected Endings

Each of the sentences below has a word missing. Read each sentence. Then carefully read each of the words below the sentence. Draw a circle around the word that best completes the sentence. Sometimes, "none of these" may be the best answer, because a correct word is not given.

A.	Her piece of cake is _____ than mine.			
	big	biggest	bigger	none of these
B.	Doctors were _____ at the hospital.			
	need	needed	needing	none of these
C.	The puppy _____ out of the box.			
	jump	jumper	jumping	none of these

Figure 1. Directions and practice items from the first page of the Inflected Endings Subtest.

The first step in constructing the Contractions & Possessives Subtest was to conduct a survey of basal materials, identifying the contractions typically taught to second, third, and fourth grade students, and the formats used to teach and test these contractions. The four widely used basal series selected for review were: Ginn 720 (Clymer et al., 1978 Rainbow Edition), Heath and Company (Witty, Bebell, & Freeland, 1968 Edition), American Book Company (Johnson et al., 1968 Edition), and Houghton Mifflin (Durr et al., 1974-78 Edition). A skills management system, the Wisconsin Design for Reading Skill Development (Otto et al., 1972, 1975), was also reviewed. The review of these materials revealed that many contractions are taught by the end of the second grade, and that all of the common contractions receive instructional attention by the end of third grade.

The second step in developing the Contractions & Possessives Subtest was to select the contractions to test and decide the number of items needed to assess each target contraction. As with the other two structure subtests, the number of target items for each contraction was based on frequency information. First, the contractions were grouped into categories based on which member of the word pair was contracted. For example, contractions of will, such as I'll, we'll, he'll, and they'll formed one contraction category. Then, The American Heritage Word Frequency Book[1]

---

[1] This book is a word list based on the examination of published material for children in third through ninth grade, and contains 5,088,721 tokens and 86,741 words. Over 5 million words of running text were extracted for analysis from 1,045 different publications, with a total of 90 schools participating in the study. Table 3 presents the results of the analysis of the 600 most frequent words for the presence of target inflected endings.

(Carroll, Davies, & Richman, 1971) was used to determine the frequencies of each of the specific contracted forms within the categories. Based on these frequency tabulations, the contraction categories were rank-ordered and a proportional number of specific contracted forms were selected for assessment in the new subtest. This category and frequency information is summarized in Table 3.

In addition to 21 items assessing contractions, 10 items were created to assess possessives for a total of 31 items on the Contractions & Possessives Subtest. The target possessive and contracted forms assessed in the subtest, as well as the number of items for each form, are presented in Table 4.

The format of the Contractions & Possessives Subtest required that pupils read a sentence which contained an underlined contracted or possessive word, and then circle the response choice (from four possible responses provided below the sentence) that best described the meaning of the underlined word in the sentence. Sentences for each item were brief and similar in vocabulary level to minimize variation in performance due to differences in pupils' vocabulary knowledge. Figure 2 is a copy of the directions and practice items from the Contractions & Possessives Subtest.

As mentioned above, there were four response choices for each item. The response choices for items varied, depending on whether the item assessed a contracted form not ending in apostrophe s (such as they've), a contracted form ending in apostrophe s (such as what's), or a possessive form ending in apostrophe s (such as Tony's).

Table 3

Frequency of Contractions[a] and Rank Order and Number of Test Items per Contraction Category

Contraction	Contraction frequency	Rank order by category	Number of test items by category
aren't	239	1	6
doesn't	590		
hasn't	98		
shouldn't	90		
weren't	176		
won't	756		
it's	2,178	2	3
here's	118		
what's	482		
it'll	65	3	3
they'll	120		
you'll	524		
they've	53	4	2
you've	317		
I'd (had)	534	5	2
we'd (had)	not listed		
she'd (would)	130	6	2
it'd (would)	11		
I'm	1,848	7	1
let's	892	8	1
you're	848	9	1

[a] American Heritage Word Frequency Book

Table 4

Target Forms in Winter 79/80

Contractions & Possessives Subtest

Target Forms		Occurrences
n't	(not)	6
'll	(will)	3
'd	(would)	2
've	(have)	2
'd	(had)	2
'm	(am)	1
's	(us)	1
're	(are)	1
's	(is)	3
's	(possessives)	10
Total		31

Contractions and Possessives

Read each sentence below carefully. Then decide which meaning the apostrophe mark (') has in the underlined word. Circle the choice given below the sentence that tells the meaning of the underlined word. In some cases, "none of these" may be the correct answer because the real meaning of the apostrophe in the underlined word is not given.

A.	The cuckoo clock's been broken for a long time.			
	possessive '	clock is	clock has	none of these
B.	There wasn't any question that Andy was the best runner.			
	possessive '	was none	was no	none of these
C.	The wind's energy is used by windmills to raise water.			
	possessive '	wind has	wind is	none of these

Figure 2. Directions and practice items from the first page of the Contractions & Possessives Subtest.

The four response choices for all the items which assess a contracted form not ending in apostrophe s consisted of the correct response choice, two response choices consisting of two-word combinations which could make sense in the context of the sentence but did not correspond to the contracted form, and the phrase none of these. Thus, if the target contracted form was they've, the incorrect response choices might be "they did," "they gave," and the phrase none of these.

The formation of response choices for target words ending in apostrophe s was the same regardless of whether the apostrophe s represented a possessive or a contraction. The four response choices included the word possessive ', two response choices consisting of one word of the two-word phrase which comprised the target contracted form, and the phrase none of these. For example, for the target word what's, the response choices were (1) possessive ', (2) what has, (3) what is, and (4) none of these. For the target word Jan's, the response choices were (1) possessive ', (2) Jan has, (3) Jan was, and (4) none of these. The only exception in the formation of the response choices was when the phrase none of these was the correct answer.

Affixes Subtest

A 54-item subtest was developed to assess students' knowledge of affixed words. A total of 18 target affixes was assessed in the revised subtest, compared to 20 assessed in the previous version. The target affixes that were selected for the subtest, as well as the number of items for each type of affix, are presented in Table 5.

Table 5

Comparison of Target Affixes Between

Spring 1979 and Winter 79/80 Subtests

Target Affixes	Spring 1979 Number of Test Items	Winter 79/80 Number of Test Items
dis-	4	3
in-/im-	4	3
inter-	4	3
non-	4	3
pre-	4	3
re-	4	3
sub-	4	3
un-	4	3
-able/-ible	4	3
-ance/-ence	4	-
-ant/-ent	4	-
-en	4	3
-er/-or	4	3
-ful	4	3
-ion/-tion/-sion	4	3
-less	4	3
-ly	4	3
-ment	4	3
-ness	4	3
-ous	4	3
Total	80	54

Analysis of the data from the previous administration of the Affixes Subtest had raised several issues regarding the design of the test items and the length of the subtest. These concerns resulted in four kinds of revisions on the Affixes Subtest.

The first type of revision concerned the selection of root words used in assessing the 18 target affixes. The previous version of the subtest included several root words containing bound bases.[2] There was some concern that words with bound bases might confound the assessment of children's knowledge of the affixed element. Therefore, despite the high frequency of affixed words containing bound bases in the English language, it was decided that the number of root words with bound bases in the revised Affixes Subtest would be limited. When words containing bound bases are used in the revised subtest, the bound form of the word occurs in all four response choices. For example, in the item with the root word collect, which contains the bound base lect, the response choices are all extensions of the word collect (i.e., collection, recollect, collector, collectible). Lecture or elect do not qualify as response choices. Similarly, in the item using the root word invent, which contains the bound base vent, all response choices include invent; thus prevent does not qualify as a distractor.

The second area of revisions focused on the vocabulary level of the root words and their corresponding affixed forms. Elementary school

---

[2] A bound base is a linguistic element, rather than an independent, meaningful word. An example of a bound base is the element lect as in the words collect and lecture. Approximately half of all suffixes and 73% of all prefixes attach to bound bases.

children's familiarity with each root and affixed word in the previous version of the Affixes Subtest was determined from information in The Living Word Vocabulary (Dale & O'Rourke, 1976).[3] For many items in the subtest, the root word was familiar to students at a particular grade level, but the affixed forms of the same root word were not. For example, the root word act is listed in The Living Word Vocabulary as being familiar to a majority of second, third, and fourth grade pupils. Yet several of the response choices, affixed forms of the root word act such as inactive or activated, are familiar only to pupils at the eighth grade level and above. To minimize the variations in familiarity ratings between a root word and its affixed forms, response choices were selected that had vocabulary levels as consistent as possible both within and across test items. Root words from the previous version of the subtest which did not yield affixed forms familiar to second, third, and fourth grade pupils were replaced with root words whose affixed forms are well known by students at these grade levels.

A third area of concern regarding the previous version of the Affixes Subtest was the variation in spelling between certain root words and their affixed forms. For example, in the previous subtest, the root word day is changed to daily when the suffix -ly is added. To minimize the effect of such alterations in spelling, the revised version of the Affixes Subtest uses the complete affixed forms of the root word as response choices rather than the isolated affixes used in the previous version. The

---

[3] The Living Word Vocabulary lists 43,000 words and their percentage scores based on how familiar the words are to students in grades 4, 6, 8, 10, 12, 13, and 16.

affixed portions of each response choice are also underlined in the revised subtest to help students focus their attention on the affixed component. Furthermore, the base or root word for each item is shown beneath the item number and is set off in a box for easy recognition.

Finally, a decision was made to shorten the Affixes Subtest by deleting one item from each of the 20 affix categories and to eliminate two of the target affixes, -ant/-ent and -ance/-ence. Performance had been lower on these two affixes than on the other affixes assessed in the previous version. Moreover, these two affixes created problems in item construction because their affixed forms occur mainly in words rated far above the appropriate vocabulary level for the subtest. The present version of the Affixes Subtest consists of only 54 items, and requires a shorter administration time. This adjustment in length helps to make the Affixes Subtest more appropriate for second and third grade students.

The format for the revised Affixes Subtest is similar, but not identical, to the format of the previous version. In both tests, pupils read a prose description of the affixed word and were instructed to circle the response choice (from four response choices provided below the description) defined in the stem. In the revised subtest, however, the stems are shortened and the root word is shown beneath each item number. The response choices for the revised subtest include the correct response affixed word and three other affixed forms of the same root word which do not fit the description or definition given in the item stem. In the previous version of the subtest, response choices consisted of only the isolated

affixes. Figure 3 is a copy of the directions and practice items from the current Affixes Subtest.

PROCEDURE

Each classroom of children received all three Structure Subtests: Inflected Endings, Contractions & Possessives, and Affixes. The Inflected Endings Subtest was given first. Administration time for the Inflected Endings Subtest (including directions and practice items) was approximately 25 minutes for second grade students and 15 minutes for third grade students. After a 30-minute break for recess, the Contractions & Possessives Subtest was administered; this subtest (including directions and practice items) took approximately 38 minutes for second grade students and 35 minutes for third grade students. Total administration time for the Affixes Subtest, the last subtest given, was approximately 33 minutes for second grade students and 25 minutes for third grade students.

All tests were administered by staff members of the Wisconsin Research and Development Center for Individualized Schooling with the assistance of one additional person hired by the Project to aid with test administration. Testing was conducted one morning in December 1979.

Results

STRUCTURE SUBTESTS

Summary statistics for the 80 second and third grade students on the Structure Subtests (Inflected Endings, Contractions & Possessives, and Affixes) are presented in Table 6.

Affixes

Look at each row and read the word in the small box. This word is the root or base word for the sentence. Now read the sentence. Below the sentence are four answers, each containing the root word plus another word part or parts. Draw a circle around the word that is described or defined in the sentence.

A. happy	A word that describes a person who is not happy: unhappy happiest happily unhappiness
B. sweet	A word that means the quality of being sweet: unsweetened sweetness sweeter sweetest
C. drive	A word that describes a car that can not be driven: driving driver drivable undrivable

Figure 3. Directions and practice items from the first page of the Affixes Subtest.

Table 6

Summary Statistics for the

Structure Subtests

($\underline{N}$ = 80)

Subtest	Number of items	$\bar{X}$ percent correct	S.D.	Range (percent correct)
Inflected Endings	39	79.23	18.31	23.08 - 97.44
Contractions & Possessives	31	58.95	20.65	19.35 - 96.77
Affixes	53	48.61	22.13	0.00 - 90.57

The large variance in mean scores between the three subtests led to certain conclusions regarding the grade level appropriateness of these subtests. This issue is discussed in the concluding section of this report.

Data from the three Structure Subtests were also used in t-tests for significant differences in performance due to sex and grade level. Summary information on the t-tests for sex and grade level differences is presented in Tables 7 and 8, respectively.

Because percentile scores from a standardized reading test were made available by the participating school, another analysis was performed on the Structure Subtests data. The total sample was stratified into high, medium, and low reading groups based on subjects' percentile scores on the standardized reading tests. T-tests were used to examine between-group differences on each structure subtest and on the structure component as a whole. Summary information from this analysis is presented in Table 9.

Students' performance on each of the three Structure Subtests was also examined by item categories within each subtest. This information, as well as reliability estimates for each subtest, is presented below.

Inflected Endings Subtest

Of the 39 items on the Inflected Endings Subtest, students obtained a mean score of 30.89 (79.23%) correct. In order to identify items on which students performed poorly, and to obtain an estimate of the reliability of the Inflected Endings Subtest as a whole, the raw score data

Table 7

Summary of t-Tests for Differences Due to Sex

(N = 80; df = 78)

	N	$\bar{X}$	t-value	Probability
Inflected Endings				
Boys	42	.75	2.11	.038
Girls	38	.84		
Contractions & Possessives				
Boys	42	.55	1.94	.056
Girls	38	.64		
Affixes				
Boys	42	.47	.72	.47
Girls	38	.51		
Total Test				
Boys	42	.58	1.58	.118
Girls	38	.64		

Table 8

Summary of t-Tests for Differences

Due to Grade Level

(N = 80; df = 78)

	N	$\bar{X}$	t-value	Probability
Inflected Endings				
Grade 2	43	.73	3.83	.000
Grade 3	37	.87		
Contractions & Possessives				
Grade 2	43	.54	2.53	.013
Grade 3	37	.65		
Affixes				
Grade 2	43	.39	4.49	.000
Grade 3	37	.59		
Total Test				
Grade 2	43	.53	4.31	.000
Grade 3	37	.69		

Table 9

Description of Reading Groups and Summary of

t-Tests by Subtest

	Low 0-33 percentile	Medium 34-79 percentile	High 80-99 percentile
N	13	45	22
$\bar{X}$	.4753	.5799	.7483
S.D.	.112	.182	.132
S.E.	.031	.027	.028

Comparisons

	Low-Medium (df = 56)		Medium-High (df = 65)		Low-High (df = 33)	
	t-value	Probability	t-value	Probability	t-value	Probability
Inflected Endings	1.55	.126	3.38	.001	4.85	.000
Contractions & Possessives	2.23	.030	3.63	.001	6.25	.000
Affixes	1.60	.116	3.24	.002	4.43	.000
Total Test	1.96	.055	3.86	.000	6.25	.000

were subjected to an item analysis. The resultant Hoyt reliability estimate for the subtest was .91 with a standard error of measurement of 2.06.

A rank-ordered listing of students' performance on each inflected ending category is presented in Table 10. The mean percent correct on these categories ranged from 90.8% (for the ing ending) to 51.2% (for the item assessing "kept").

Contractions & Possessives Subtest

Students obtained a mean score of 18.27 (58.95%) correct on the 31 items on the Contractions & Possessives Subtest. In order to identify items on which students performed poorly and to obtain an estimate of the reliability of the subtest, the raw score data were subjected to an item analysis. The resultant Hoyt reliability estimate for the subtest was .88 with a standard error of measurement of 2.16.

A rank-ordered listing of students' performance on each of the contraction categories and on the posssessive apostrophe category is presented in Table 11. Mean percent correct on these categories ranged from 90.0% (for the contraction am) to 36.7% (for the possessive apostrophe category).

Affixes Subtest

Of the 53 items on the Affixes Subtest, students obtained a mean score of 25.76 (48.61%) correct. In order to identify items on which students performed poorly and to obtain an estimate of the reliability of the Affixes Subtest, the raw score data were subjected to an item analysis. The resultant Hoyt reliability estimate for the subtest was .93 with a standard error of measurement of 3.11.

Table 10

Rank-ordered Listing of Mean Percent Correct on Inflected Endings Item Categories

Category	Number of Items	Mean Percent Correct
ing	3	90.8
root word (as correct response)	11	85.7
er	3	80.8
es	6	79.9
est	3	79.9
y	1	77.5
(i)ed	6	73.2
"kept" (tense marker with vowel change)	1	51.2

Table 11

Rank-Ordered Listing of Mean Percent Correct

on Contractions & Possessives Categories

Category	Number of Items	Mean Percent Correct
am	1	90.0
is	3	81.7
are	1	81.3
us	1	78.7
not	6	69.3
will	3	65.4
have	2	63.7
had	2	58.1
would	2	54.4
Possessive apostrophe	10	36.7

A rank-ordered listing of the affix categories is presented in Table 12. Mean percent correct on these categories ranged from 67.1% (for the -ful category) to 27.9% (for the -(t)ion category).

Table 12

Rank-ordered Listing of Mean Percent Correct on Affixes Categories

Category	Number of Items	Mean Percent Correct
Prefixes:	(23)	
re-	3	62.1
non-	3	53.7
dis-	3	52.0
un-	2*	48.1
in-/im-	3	43.7
sub-	3	37.1
inter-	3	32.4
pre-	3	31.2
Suffixes:	(30)	
-ful	3	67.1
-or/-er	3	58.3
-less	3	55.0
-able/-ible	3	53.7
-ment	3	52.9
-ly	3	50.8
-(i)ous	3	50.4
-en	3	47.4
-ness	3	47.1
-(t)ion	3	27.9

Note: One item assessing un- was deleted from analysis due to a typographical error in the test booklets.

DEVELOPMENT OF THE PHONICS SUBTESTS

The winter 1979/80 Phonics test consisted of two sections: a Consonants Subtest and a Vowels Subtest. The two subtests were developed independently, although similar guidelines were followed for the development of test items and for test format.

Method

SUBJECTS

The four classrooms from Belleville Elementary School participating in the study to evaluate the three Structure Subtests also participated in the study to evaluate the two Phonics Subtests. A total of 39 second grade students and 34 third grade students took both the Consonants and Vowels Subtests.

PROCEDURE

The Consonants and the Vowels Subtests were administered the same morning. After the Consonants Subtest was administered, the children had a half-hour recess, followed by the administration of the Vowels Subtest. Both subtests were given in January 1980 by staff members from the Wisconsin Research and Development Center for Individualized Schooling and by specially trained personnel hired by the Project. Directions for both subtests were read from administrator's manuals.

Consonants Subtest

After the booklets for the Consonants Subtest were distributed and the student identification information entered on the cover, the examiner explained to the students that they would be listening for consonant sounds at the beginning of words. The examiner then worked with the students on four practice items. One practice item represented each of the following categories: single-letter consonants with one common sound correspondence, single-letter consonants with more than one common sound correspondence, consonant clusters, and consonant digraphs.

For each item, students were directed to look at the synthetic word in the box and to pronounce the word silently to themselves. They were told to especially note the underlined letter(s) in the word, and to decide the sound made by that underlined part. Next, the examiner read the names of the four pictures in the row. The students' task was to circle the picture whose name began with the same sound as the sound made by the underlined part of the synthetic word.

Following discussion of the four practice items (between 8 and 13 minutes), the children were led through the test. The examiner paced the children through each test item, saying, "put your finger on Row # __" and "say the made-up word to yourself." The examiner then named the four pictures in the row. This procedure was repeated for each item on the test.

When the Initial Consonants section of the Consonants Subtest was completed, the examiner explained to the students that they would now

begin listening for consonant sounds at the end of words. The examiner worked with students on the three practice items for final consonant sounds (using as examples a single-letter consonant with one common sound correspondence, a consonant cluster, and a consonant digraph). At the end of the practice period (between 4 and 10 minutes), the examiner led the children through the actual test items naming each picture as before. Total test time for both the Initial and Final Consonants sections (including practice items) ranged from 38 to 43 minutes.

Vowels Subtest

The Vowels Subtest was administered in the same way as the Consonants Subtest, except that students were instructed to listen to vowels sounds in the middle of words (51 items) and at the end of words (5 items). Practice items were provided for items in both positions.

Total test time (including practice items) ranged from 27 to 30 minutes for both the Medial and Final Vowels sections of the Vowels Subtest.

STIMULI

Consonants Subtest

The present version of the Consonants Subtest was designed to assess 45 different spelling-to-sound correspondences. Sounds selected for testing were represented by single-letter consonants, two-letter consonant clusters, and two-letter consonant digraphs. The 45 target correspondences were assessed in 90 items, as compared with the previous version, which

assessed 25 target correspondences in 81 items. The revised Consonants Subtest was designed to be more global, assessing almost twice as many correspondences as in the previous test. The two tests, however, were approximately the same length because two, rather than three or four, items were created for each correspondence. It was intended that teachers would evaluate children's performance in the present version of the Consonants Subtest by category (i.e., single-letter consonants, consonant clusters, and consonant digraphs), rather than by individual correspondences, as was the case in the previous version of the test. For example, a student making errors on correspondences represented by cl-, -nt, -mp, gr-, and st-, would be viewed as having difficulty with consonant clusters in general, rather than with these five correspondences in particular.

Of the 90 items comprising the Consonants Subtest, 59 items assessed spelling-to-sound correspondences of consonants in initial position and 31 items assessed correspondences in final position. Each item on the subtest consisted of a synthetic word with the target letter(s) underlined and four response choices in picture form. Students were directed to read the synthetic word silently to themselves and to determine the sound represented by the underlined letter(s). Next, they were told to listen as the examiner read the names of the four picture response choices for that item. Students were then instructed to draw a circle around the picture whose name began (or ended) with the sound represented by the underlined letter(s) in the target synthetic word. Figure 4 is a copy of the directions and practice items for consonant correspondences in initial position from the Consonants Subtest.

PHONICS: Consonants
Initial Position

In each row, look at the made-up word. Notice that there is a letter or letters underlined at the beginning of that word. Read the word to yourself and decide how the underlined letter or letters sound. Then listen carefully as I read the names of the pictures in the row and decide which picture name begins with the same sound as the underlined letter or letters in the made-up word. Draw a circle around that picture.

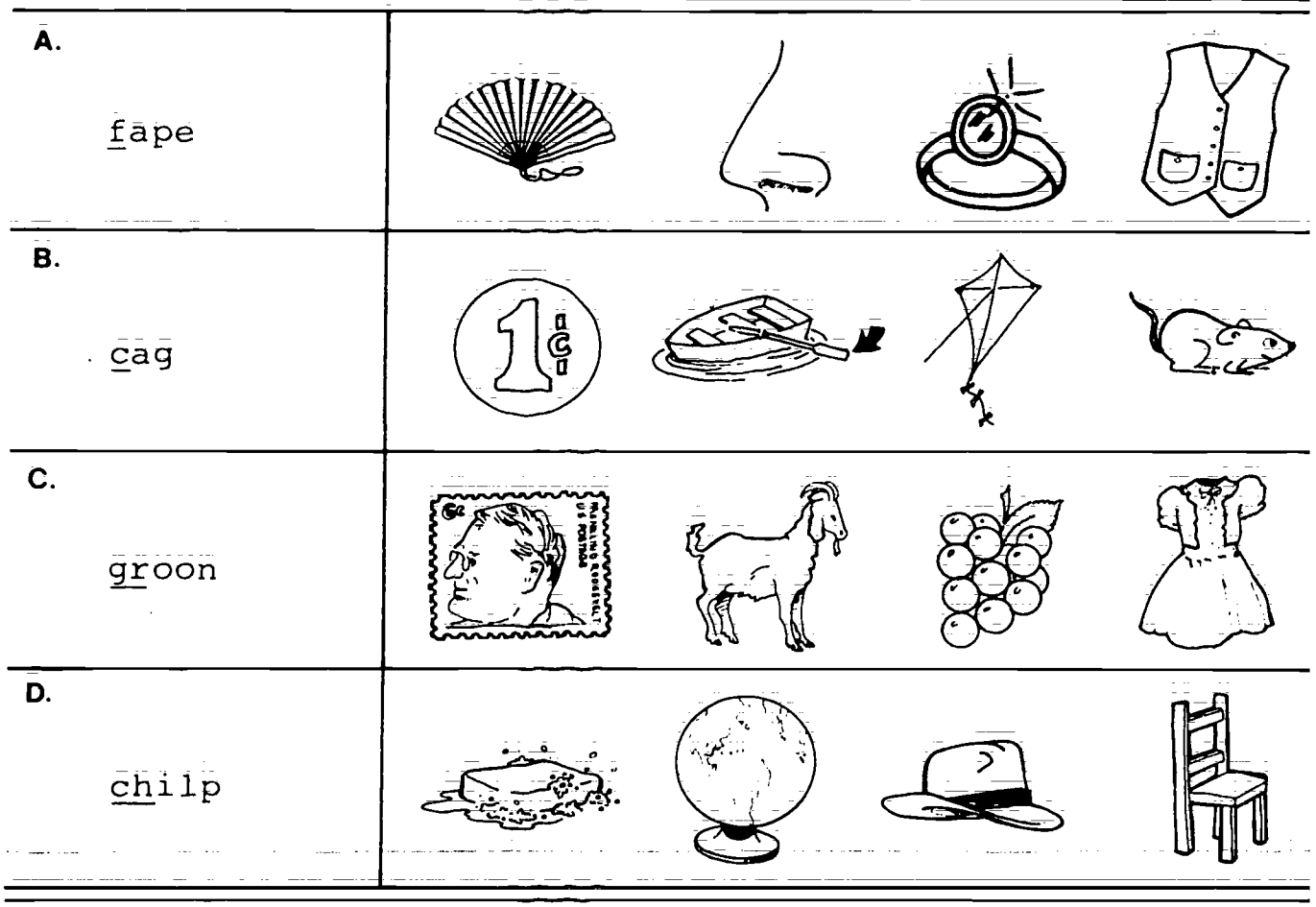

Figure 4. Directions and practice items from the first page section of the Consonants Subtest for Initial Position.

Target consonant sounds for the present test were selected according to their frequencies of occurrence in the Venezky (1962, Note 1) tabulations of the 20,000 most common English words. All single-letter consonant and two-letter consonant digraph correspondences with frequencies exceeding 150, and all two-letter consonant cluster correspondences with frequencies exceeding 110, were selected for inclusion in the Consonants Subtest. These 45 spelling-to-sound correspondences each appeared in two one-syllable target synthetic words that conformed to phonological principles of the English language.

The response choices for each item were four pictures whose names are well known to elementary school children. For each target item, the four response choice categories were: a Correct response choice, a Visually Close response choice, an Acoustically Close response choice, and a "Neither" (neither acoustically close nor visually close) response choice. For the six items assessing the three single-letter consonants having other common sound correspondences--c, g, and s--the Acoustically Close category was changed to be the Other Common Sound Correspondence.

SELECTION OF TARGET SOUNDS. Spelling-to-sound correspondences for consonants can be classified into three major categories: single-letter consonants, two-letter consonant clusters, and consonant digraphs. Of the 90 items comprising the Consonants Subtest, 59 items assessed correspondences in initial position and 31 assessed them in final position.

According to the Venezky (1962) tabulations, 23 of the 24 single-letter consonant correspondences have frequencies exceeding 150 in the 20,000 most common English words. (The exception is y which, when it

assumes consonant status in initial position, has a frequency of only 50; because of this low frequency value, y is not assessed in the Consonants Subtest). Two other single-letter consonants, x and z, were also not assessed in the Consonants Subtest. Although their frequencies exceeded 150, it was either extremely difficult or impossible to develop items for these two correspondences utilizing only one-syllable words. The 21 single-letter consonants that were selected for inclusion in the Consonants Subtest, their frequencies of occurrence, and the positions in which they were assessed are presented in Table 13.

Nineteen two-letter consonant clusters were identified as having frequencies of over 110. The 19 different consonant clusters, their frequencies of occurrence and the positions in which they were assessed (arranged by manner) are also presented in Table 13.

Of the seven two-letter consonant digraphs, only five had great enough frequencies to warrant inclusion in the Consonants Subtest. These five digraphs, all with frequencies exceeding 150, are: sh, ng, th (voiceless), ch, and ph. The frequency of occurrence of each digraph, as well as the position in which each digraph is assessed, is presented in Table 13.

SELECTION OF RESPONSE CHOICES. There were four types of response choices for each item: a Correct response choice, a Visually Close response choice, an Acoustically Close response choice, and a "Neither" (neither acoustically close nor visually close) response choice. For

Table 13

Target Correspondences in Consonants Subtest

by Frequency and Position

	Example	Total frequency	Position assessed	
			Initial	Final
	Single-letter consonants			
b	bat	1,445	XX	
c (as /k/)	cup	2,433	XX	
c (as /s/)	cent	719	XX	
d	dog	2,897	X	X
f	fox	1,064	X	X
g (as /g/)	goat	722	X	X
g (as /j/ with silent e)	cage	595		XX
h	house	764	XX	
j	jug	214	XX	
k	kite	395	X	X
l	lamp	3,679	XX	
m	mice	2,711	X	X
n	sun	4,599		XX
p	pet	1,811	X	X
q (followed by u)	queen	192	XX	
r	rake	971	XX	
s (as /s/)	sink	2,171	X	X
s (as /z/)	boys	612		XX
t	tag	4,040	X	X
v (with silent e in final position)	vest	1,534	X	X
w	witch	442	XX	

Table 13 (continued)

	Example	Total frequency	Position assessed	
			Initial	Final
Consonant clusters				
stop + liquid				
br	broom	232	XX	
cl	clip	184	XX	
cr	crown	241	XX	
dr	drum	136	XX	
gr	grapes	275	XX	
pl	plug	175	XX	
pr	prince	549	XX	
tr	train	401	XX	
liquid + stop				
lt	belt	155		XX
fricative + liquid				
fl	flag	160	XX	
fr	frog	125	XX	
sl	sled	114	XX	
nasal + stop				
mp	lamp	274		XX
nd	hand	626		XX
nt	tent	1,304		XX
nasal + fricative				
nc (followed by silent e)	fence	506		XX

Table 13 (continued)

	Example	Total frequency	Position assessed	
			Initial	Final
Consonant clusters				
fricative + stop				
sc	scale	162	XX	
sp	spool	325	XX	
st	stamp	1,054	X	X
Consonant digraphs				
ch	chair	270	XX	
ng	swing	401		XX
ph	phone	188	X	X
sh	ship	427	X	X
th (voiceless)	wreath	329	X	X

the six items assessing the three single-letter consonants having other common sound correspondences--c, g, and s--the Acoustically Close category was again replaced with a response using these Other Common Sound Correspondences.

Correct Response Choice. This response choice is a picture whose name contains an initial or final consonant sound identical to the sound corresponding to the underlined letter(s) in the target synthetic word.

Visually Close Response Choice. This response choice is a picture whose name contains an initial or final letter (or cluster or digraph) that is visually close (but not identical) to the underlined letter(s) in the target synthetic word. Data from Bouma's (1971) confusion matrix for lower case letters were used to create the Visually Close response choices.[4] For example, the Bouma matrix shows that b is most often confused as the lower case letter h. Therefore a picture whose name begins with h would be used as the visual distractor for an item assessing initial consonant b.

For many of the items across all three consonants categories, the confusions listed in the Bouma matrix were not used. In these cases, distinctive features information derived from Bouma (1971) formed the basis for selecting a Visually Close response choice. Bouma's distinctive features analysis takes into account the height of the letters

---

[4] The Bouma study reported confusions made for 25 of the 26 lower case letters when presented on a screen in IBM Courier 10 typeface, seven degrees from the point of fixation at a distance of 50 centimeters.

(short letters, ascenders, and descenders), and the width and contour of the letters (slender letters, rectangular envelopes, round envelopes, and triangular envelopes). Letters sharing the same properties are considered to be visually similar.

The Bouma confusion matrix and distinctive features information was not applicable for two-letter consonant clusters and consonant digraphs. For these correspondences, the Visually Close response choice was a picture whose name began (or ended) with only one of the two letters of the cluster or digraph. This foil could differentiate between children who attend to the complete cluster or digraph (two letters) and those who attend to only a portion (one letter) of the cluster or digraph. In an item assessing the digraph ph in final position, for example, the Visually Close reponse choice was a picture of a map, because map ends with p, one of the two letters of the digraph ph. It should be noted that because the Visually Close reponse choice for clusters contains one of the two letters of the cluster, this response choice is also acoustically close to the target cluster.

Acoustically Close Response Choice. This response choice, used for all the consonant correspondences except the three single-letter consonants which have two common sound correspondences, is a picture whose name contains an initial or final consonant whose sound is close, but not identical, to the sound of the underlined letter(s) in the target synthetic word. Whenever feasible, data from the Miller and Nicely (1955) confusion

matrices were used for selecting an acoustically close sound for single-letter consonant items.[5] For example, the Miller and Nicely matrices indicate that /b/ is most often confused as either /v/ or /f/. Therefore, a picture whose name begins with /v/ or /f/ might be used as the acoustic distractor for an item assessing the initial consonant /b/.

The Miller and Nicely information was not useful for some of the single-letter consonants or for any of the consonant clusters or consonant digraphs. The Acoustically Close response choice for those single-letter consonants for which Miller and Nicely data could not be used was based upon speech data describing where the various consonant sounds are produced within the vocal tract. Two sounds produced in proximal loci within the vocal tract are perceptually closer than two sounds produced farther away from each other. Acoustically close response choices for some single-letter consonants and for all of the consonant digraphs were formulated on the basis of this production-perception information. For example, the selection of an Acoustically Close response choice in initial position for /dʒ/ (as in jar) was accomplished by identifying a sound produced in a proximal locus within the vocal tract to where /dʒ/ is produced. The sound /tʃ/ (as in chair) fulfills this requirement and, therefore a picture of a chair could be an Acoustically Close response choice.

---

[5] The Miller and Nicely study presented perceptual confusions for 16 consonant singletons under the manipulated variables of signal to noise ratio, decibel level, and frequency response in cycles per second.

A separate set of decision rules determined the Acoustically Close response choices for consonant clusters. Acoustically Close response choices for clusters were developed by changing the place (i.e., front to back in the vocal tract), but retaining the manner (e.g., stop, liquid, fricative) of one of the two letters of the cluster. For example, an Acoustically Close response choice for the target cluster /gr/ could be /dr/ or /br/ (the /g/ is changed to another voiced stop, but the liquid /r/, or even /gl/, is retained). Table 14 lists the types of changes made in target consonant clusters for the Acoustically Close response category. For the three single-letter consonants which have two common sound correspondences (c, g, and s), the Acoustically Close response choice was replaced by a new category, the Other Common Sound Correspondence. Responses in this category were pictures whose names began or ended with the second most frequently occurring sound correspondence for the target single-letter consonant (c as /s/, g as /j/, and s as /z/).

"Neither" Response Choice. This response choice is a picture whose name has neither a sound that is acoustically close, nor a letter that is visually close, to the underlined part of the target synthetic word. For example, in an item assessing the single-letter consonant h in the target synthetic word hurp, the "Neither" response choice was a picture of a sock. The sound /s/ in sock is neither acoustically close to the /h/ in hurp, nor is the letter s visually close to the letter h. For consonant clusters and consonant digraphs, the "Neither" response choice began or ended with a two-letter consonant cluster that was neither acoustically close to the sound,

Table 14

Categories for Determining Acoustically Close

Response Choices for Consonant Clusters

Consonant Clusters		Change to another stop	Keep stop	Change to another liquid	Change to another fricative	Keep fricative	Change to another nasal	Keep nasal	Keep nasal	Keep fricative
		Keep liquid	Change to another liquid	Change to another stop	Keep liquid	Change to another liquid	Change to another stop	Change stop	Change voiceless fricative to voice fricative	Change to another stop
Initial	Final									
stop + liquid										
br-		gr, dr								
cl-		pl	cr							
cr-		pr	cl							
dr-		gr, br								
gr-		dr, br								
pl-		cl (/kl/)	pr							
pr-		cr (/kr/), tr								
tr-		pr, cr (/kr/)								
	-lt	lk		rk						
fricative + liquid										
fl-					sl	fr				
fr-					thr	fl				
sl-					fl (2)					
nasal + stop										
	-mp						nt, nk			
	-nd						nk	nt		
	-nt							nk, nd		
nasal + fricative										
	-nce								nz	
fricative + stop										
sc-										sp, st
sp-										st (2)
st	-st									sp, sc (/sk/)

nor visually close to the letters of the target consonant cluster or digraph. The Neither category was created to see whether children, when they were unsure of a correct answer, would be strategic and tend to select either an Acoustically Close or a Visually Close response, or whether they would merely guess at the correct answer randomly selecting from among all three incorrect response choices. If error analyses of individual children's performance revealed as many responses in the Neither category as in the Acoustically Close and Visually Close categories, these children would be using random rather than strategic selection processes.

Vowels Subtest

The new version of the Vowels Subtest was designed to assess 28 different spelling-to-sound correspondences. Sounds selected for testing included the five short and five long vowels, four other frequently occurring single-letter vowels (i.e., single-letter vowels that correspond to sounds that are neither short nor long), and 14 two-letter vowel clusters. The 28 spelling-to-sound correspondences were assessed in 56 items, as compared with the previous version of the test, which assessed 20 different correspondences in 60 items. The revised test was designed to be more global, assessing 40% more correspondences in a test of approximately the same length. This was accomplished by creating two, rather than three, items for each correspondence. As with the Consonants Subtest, it was intended that teachers would analyze children's performance by category (short vowels, long vowels, other single-letter vowels, and two-letter

vowel clusters), rather than by individual correspondences. For example, a student making errors on correspondences represented by ou, ai, oo, and ow would be viewed as having difficulty with vowel clusters in general, rather than with these four correspondences in particular.

Each of the 56 items on the Vowels Subtest consisted of a synthetic word with the target letter(s) underlined, and four response choices in picture form. Students were directed to read the synthetic word silently to themselves and to determine the sound represented by the underlined letter(s). Next, they were told to listen as the examiner read the names of the four picture response choices for that item. Students were then instructed to draw a circle around the picture whose name contained the same medial (or final) vowel sound as the sound of the underlined letter(s) in the target synthetic word. Figure 5 is a copy of the directions and practice items from the Medial Vowels section of the Vowels Subtest.

Target vowel sounds were selected according to their frequencies of occurrence in the Venezky (1962) corpus of the 20,000 most common English words. The frequency ranges for the vowels selected for inclusion in the test, by category, were: short vowels = 7554 to 1458; long vowels = 1870 to 503; other single-letter vowels = 243 to 117; and two-letter vowel clusters = 723 to 123.[6] Each spelling-to-sound correspondence selected

[6]The frequency value of 723 represents the summed frequencies of 333 for or (as in porch) and 390 for or (as in corn). Although Venezky differentiates between the two correspondences, they are regarded as one correspondence in the present study.

PHONICS: Vowels
Medial Position

In each row, look at the made-up word. Notice that there is a letter or letters underlined in the middle of that word. Read the word to yourself and decide how that letter or letters sound. Then listen carefully as I read the names of the pictures in the row and decide which picture name contains the sound of the underlined letter or letters in the made-up word. Draw a circle around that picture.

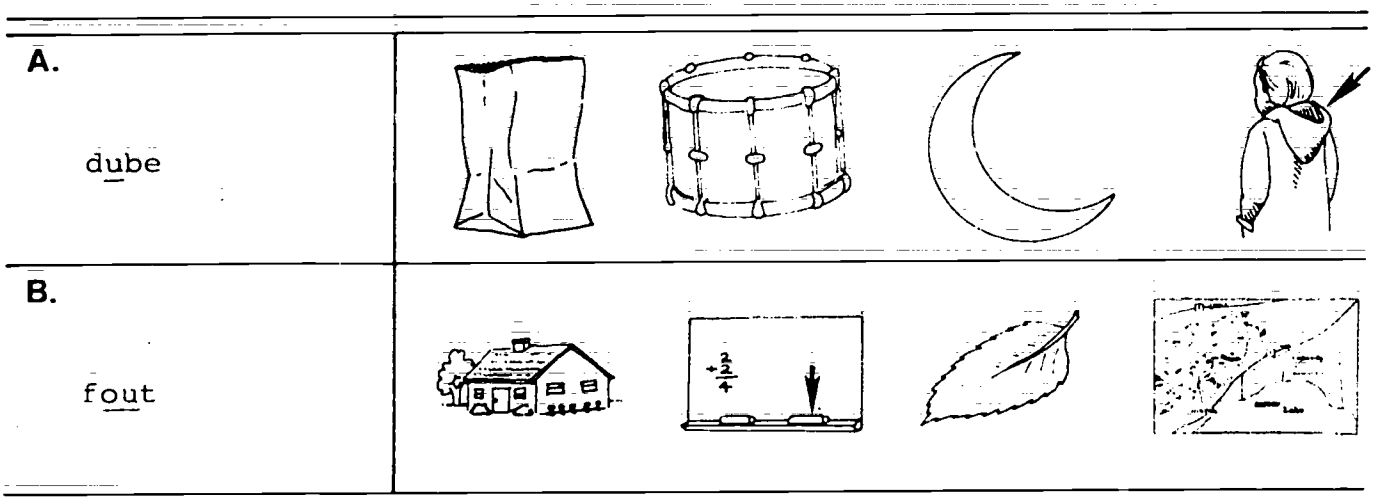

Figure 5. Directions and practice items from the first page of the Vowels Subtest for Medial Position.

appeared in two one-syllable target synthetic words that conformed to phonological principles of the English language.

The response choices for each item were four pictures whose names are well known to elementary school children. For each target item, the four response choice categories were: a Correct choice, an Acoustically Close response choice, a Visually Close response choice, and a Neither (neither acoustically close nor visually close) response choice.

Of the 56 items in the Vowels Subtest, 51 items assessed vowels in medial position in words, and 5 items assessed vowels in final position. All 51 of the items testing vowel sounds in medial position used one-syllable target synthetic words and one-syllable picture names which conformed to a CVC (or CCVC) pattern. Because the final consonant sound in a CVC (or CCVC) word affects the duration of the medial vowel (and, in 51 of the Vowels Subtest items, the medial vowel was the sound of interest), either voiced or all voiceless final consonants were used for the final consonant sounds for the target synthetic word and the four response picture names within an item. This control provided some degree of consistency for medial vowel duration within items.

SELECTION OF TARGET SOUNDS. The correspondences assessed in the Vowels Subtest can be grouped into four categories: short vowels, long vowels, other common correspondences for single-letter vowels, and two-letter vowel clusters.

The five single-letter vowels corresponding to short vowel sounds were selected for testing. According to Venezky (1962), the five short

vowel sounds--short a (as in bat), short e (as in peg), short i (as in lid), short o (as in shot), and short u (as in duck)--have letter-to-sound correspondences with frequencies exceeding 150. All five short vowel sounds are designated as simple vowels. The frequencies of occurrence of each short vowel sound and the position(s) within words in which each vowel is assessed are presented in Table 15.

All five single-letter vowels corresponding to long vowel sounds were selected for testing. The Venezky (1962) tabulations indicate that the five long vowel sounds--long a (as in cape), long e (as in mete), long i (as in hide), long o (as in bone), and long u (as in flute)[7]--all have spelling-to-sound correspondences with frequencies exceeding 150. According to Kenyon and Knott's (1953) "A Pronouncing Dictionary of American English" (the source used by Venezky for the pronunciations of the 20,000 most common English words), long e and long u are always designated as simple vowels, whereas long i is always designated as a diphthong. In this study, both long a and long o were considered diphthongs because the vowel sounds in CVC trigrams are more commonly pronounced as diphthongs than as simple vowels. In this study, all target synthetic words and

---

[7] While the sound of u in fuse (the diphthong ju or iu) has a greater frequency than the sound of u in flute (the simple vowel u), only the latter letter-sound correspondence was used in the Vowels Subtest. This is because there are few one-syllable picturable words with u as in fuse that are well-known to elementary school pupils. Except for when the sound correspondence for the letter u occurs in initial position, the two sound correspondences above for u are considered very similar.

Table 15

Target Correspondences for Short, Long, and "Other"

Single-letter Vowels by Frequency and Position

	Example	Total frequency	Position assessed	
			Medial	Final
Short vowels				
a	hat	3,121	XX	
e	dress	3,241	XX	
i	fish	7,554	XX	
o	mop	1,590	XX	
u	drum	1,458	XX	
Long vowels				
a	rake	1,870	XX	
e	mete	503	XX	
i	hive	968	XX	
o	rope	1,292	XX	
u[a]	flute	967	XX	
Other single-letter vowels				
a	ball	147	XX	
o	glove	159	XX	
o	dog	117	XX	
y (sometimes followed by silent e)	fly	243		XX

[a]See Footnote 7.

picture names of response choices used in assessing vowel sounds in medial position are CVC (or CCVC) trigrams. The frequency of occurrence of each long vowel sound and the position in which each vowel is assessed are presented in Table 15.

Four other single-letter vowel sounds which did not correspond to either short or long vowel sounds were assessed in the revised Vowels Subtest. All four of these correspondences had frequencies exceeding 115 in the Venezky (1962) corpus and appeared in words well known to elementary school age children. Three of these correspondences have sounds designated as simple vowels: a (as in walk), o (as in glove), and o (as in dog). The fourth correspondence, y (as in type), has the sound of long i, and was thus designated as a diphthong. The frequency of occurrence for these four other single-letter vowel sounds and the position within words in which each correspondence is assessed are presented in Table 15.

The revised version of the Vowels Subtest assesses 14 two-letter vowel cluster correspondences, as compared to the previous version, which assessed only 10 of the vowel clusters. The additional clusters included for assessment were four of five possible vowel + r correspondences: ar (as in barn), er (as in herd), or (as in porch, and as in corn)[8], and ur (as in nurse)[9]. (Because ir [as in bird], which is the fifth vowel + r

---

[8]Venezky (1962) differentiates between the or in porch and the or in corn. In this paper, the two correspondences will be treated as one.

[9]In this subtest, all vowel + r combinations are treated as vowel clusters. The authors are aware, however, that er and ur are simple vowels, whereas ar and or are vowel + r combinations.

correspondence, had a frequency of only 106, it was not selected for inclusion in the Vowels Subtest.) Table 16 presents the frequency of occurrence for each of the 14 vowel clusters and the position in which each cluster is assessed.

SELECTION OF RESPONSE CHOICES. As in the Consonants Subtest, there were four categories of response choices for each item: a Correct response choice, a Visually Close response choice, an Acoustically Close response choice, and a Neither (neither acoustically close nor visually close) response choice.

Correct Response Choice. This response choice is a picture whose name contains a medial or final vowel sound identical to the sound represented by the underlined letter(s) in the target synthetic word.

Visually Close Response Choice. This response choice is a picture whose name contains a medial or final letter (or letters) that is either visually close or identical to the underlined letter(s) in the target synthetic word. Because the Bouma (1971) study indicated that all lower case vowels are visually similar (except for lower case i), confusion matrix information was not as useful in creating Visually Close response choices for the Vowels Subtest as it had been for the Consonants Subtest. Therefore, a new set of decision rules was created for this response category.

For the two-letter vowel clusters, the Visually Close response choice either contained one of the two letters of the cluster (e.g., for the ou cluster in proun, the Visually Close response choice was knob--only the o of the ou cluster was retained), or kept one letter constant, but changed

Table 16

Target Correspondences for Vowel Clusters by Frequency and Position

	Example	Total frequency	Position assessed	
			Medial	Final
Vowel clusters				
ai	train	261	XX	
au	taut	175	XX	
ar	barn	532	XX	
ay	play	142		XX
ea	seal	320	XX	
ea	bread	135	XX	
ee	feet	294	XX	
er[a]	fern	387	XX	
oo	moor	198	XX	
or[b]	porch	723	XX	
ou	cloud	238	XX	
ow	gown	123	XX	
ow	snow	130	X	X
ur	purse	204	XX	

[a] See Footnote 9.

[b] See Footnote 8.

the other letter of the cluster (e.g., for the ay in chay, the Visually Close response choice was saw--the a was retained in the spelling, but the y was changed to a w).

The Visually Close response category was renamed Other Frequent Sound Correspondence for the items assessing all short, all long, and the four other frequent correspondences for single-letter vowels. For the long and short vowel correspondences, the Other Frequent Sound Correspondence consisted of the long vowel counterpart for the short vowel, or the short vowel counterpart for the long vowel.[10] For example, for the synthetic word chun, with the target short u (as in drum), the response choice in the Other Frequent Sound Correspondence category was tube--a word containing the vowel sound of long u.

For the four other single-letter vowel correspondences (which were neither short nor long), the response choice for the Other Frequent Sound Correspondence was developed in a similar way. Here the response choice consisted of a picture whose name contained the same vowel spelling, but with a different sound correspondence from the vowel in the target synthetic word. For example, for the correspondence of o as /ɔ/ (as in dog), the picture representing the Other Frequent Sound Correspondence was cob and the correct answer was saws. Although the target vowel in the synthetic

---

[10] The only exception is for the target vowel, long e. Because the short e vowel sound is acoustically close to the sound corresponding to long e, the Other Frequent Sound Correspondence category had picture names containing the letter e, but which sounded like neither short e nor long e (e.g., toes and shoes).

word is always a single-letter vowel, the vowel sound in the Other Frequent Sound Correspondence may be spelled with either one or two letters (as with /u/ in tube or suit). In either case, the vowel used in the target synthetic word is included in the spelling of the picture name in the response choice in the Other Frequent Sound Correspondence category. (This inclusion of the same vowel causes the items in the Other Frequent Sound Correspondence category to be Visually Close as well.)

Acoustically Close Response Choice. This response choice is a picture whose name contains a medial or final vowel sound that sounds similar but is not identical to the sound corresponding to the underlined letter(s) in the target synthetic word. For single-letter vowels and two-letter vowel clusters corresponding to simple vowel sounds, data from the Peterson and Barney (1952) confusion matrix were used to create the Acoustically Close response choice.[11] For example, the Peterson and Barney matrix indicated that /æ/ (as in hat) is most often confused as /ɛ/ (as in neck).

The Peterson and Barney information was not utilized, however, when single-letter vowels and two-letter vowel clusters corresponded to diphthongs. Acoustically Close response choices for the eight diphthong correspondences[12]

---

[11] The Peterson and Barney study reported auditory confusions by adult observers under an approximately 70 decibel level for 10 simple vowel sounds, each presented 152 times by men, women, and children speakers.

[12] The eight diphthong correspondences are: long a, long i, long o, ai (as in wait), ay (as in pay), ou (as in house), ow (as in gown) and ow (as in bowl).

were based upon speech data which describes where the various diphthong sounds are produced within the vocal tract. As with acoustically close consonant sounds, it is known that vowel sounds produced in proximal loci within the vocal tract are perceptually closer than sounds produced in more distal loci. For example, the vowel cluster ou, which corresponds to the diphthong /aʊ/ (as in house), is produced from a low-central /a/ to a lower-high back /ʊ/ position in the vocal tract. An acoustically close vowel sound could be /ɔ/ (as in pawn), which is produced in the higher low-back position in the vocal tract--quite close to where the /aʊ/ is produced.[13]

"Neither" Response Choice. This response choice is a picture whose name has neither a sound that is acoustically close, nor a letter that is visually close, to the underlined part of the target synthetic word. For example, in an item assessing the medial vowel for short o (as in the target synthetic word, stom), wig was used as the Neither response choice because the medial vowel i in wig neither sounds like nor looks like the o in stom.

As in the Consonants Subtest, the Neither response category was created to see whether a child who did not choose the correct response choice would be strategic and tend to choose either an Acoustically Close

---

[13]The terms and locations of the positions within the vocal tract for these vowel sounds are taken from Kenyon and Knott, "A Pronouncing Dictionary of American English," 1953, p. xiii.

or a visually close response choice--or, would merely guess at the correct answer, and randomly select from among the three incorrect response choices.

Results

Summary statistics for the 73 second and third grade students on the Phonics Subtests (Consonants and Vowels) are presented in Table 17. Overall, performance on the Consonants Subtest was high (mean percent correct, 91.22) indicating that by the second and third grades, most children have mastered a majority of the letter-sound correspondences assessed by the Consonants Subtest.

Performance on the Vowels Subtest was considerably lower (mean percent correct, 70.52). The lower performance on the Vowels Subtest was expected due to the greater variability of sounds corresponding to each vowel letter. Students' performance on the two Phonics Subtests was also examined by item category. This information, as well as reliability estimates for each subtest, is presented below.

CONSONANTS SUBTEST

Of the 90 items on the Consonants Subtest, students obtained a mean score of 82.09 (91.22%) correct. In order to identify items on which students performed poorly and to obtain an estimate of the reliability of the Consonants Subtest, the raw score data were subjected to an item analysis. The resultant Hoyt reliability estimate for the Subtest was .79, with a standard error of measurement of 2.11.

Table 17

Summary Statistics for Phonics Subtests

	N	$\bar{X}$ percent correct	Range of 90 items (percent correct)
Consonants			
Grade 2	39	89.91	74 - 98
Grade 3	34	93.10	87 - 99
Total	73	91.22	74 - 99
Vowels			
Grade 2	39	66.12	30 - 96
Grade 3	34	75.94	39 - 96
Total	73	70.53	30 - 96

A rank-ordered listing of performance on the four categories of correspondences assessed in the Consonants Subtest is presented in Table 18. Performance on these categories ranged from 97.54 mean percent correct on single-letter consonants with one common sound correspondence to 61.54 mean percent correct on variant single-letter consonants. This finding replicates the finding in the previous study: children perform best on single-letter consonants which have one common sound correspondence, and least well on variant single-letter consonants (those having more than one common sound correspondence).

VOWELS SUBTEST

On the 56 items in the Vowels Subtest, students obtained a mean score of 39.49 (70.52%) correct. In order to identify items on which students performed poorly and to obtain an estimate of the reliability of the Vowels Subtest, the raw score data were subjected to an item analysis. The resultant Hoyt reliability estimate for the Subtest was .89 with a standard error of measurement of 2.94.

A rank-ordered listing of performance on the four categories of vowel correspondences is presented in Table 19. Performance ranged from a high of 79.62 mean percent correct on Vowel Clusters to a low of 48.63 mean percent correct on items assessing other single-letter vowels (vowels that were neither short nor long). These results are consistent with the findings from the previous study: children perform best on vowel clusters and least well on single-letter vowels that are neither short nor long.

Table 18

Rank-ordered Listing of Mean Percent Correct Scores by Consonant Category

Category	Number of items	Mean percent correct
Single-letter consonants[a]	29	97.54
Consonant clusters	40	95.34
Consonant digraphs	10	89.05
Variant single-letter consonants (c, s, and g)[b]	11	61.54

[a]Single-letter consonants with one common sound correspondence.

[b]Single-letter consonants with more than one common sound correspondence.

Table 19

Rank-ordered Listing of Mean Percent Correct Scores by Vowel Category

Category	Number of items	Mean percent correct
Long vowels	11	66.76
Short vowels	11	74.34
Other single-letter vowels[a]	10	[illegible].63
Vowel clusters	24	79.62

[a]Single-letter vowels that are neither short nor long.

PREPARATION OF THE FINAL VERSION OF THE WORD IDENTIFICATION TEST BATTERY

There were two main purposes of the winter 79/80 administration of the Structure and Phonics components of the Word Identification Test battery. The first was to evaluate each subtest prior to preparation of the final versions of the subtests. Overall, only minor changes were made on the subtests; these changes are documented below. The second purpose was to evaluate the grade-level appropriateness of each subtest. The Affixes and the Contractions & Possessives Subtests, for example, were found too difficult for second grade students, but appropriate for third grade students. The Phonics Subtests and the Inflected Endings Subtest appeared to be appropriate for both second and third grade students.

Structure Subtests

INFLECTED ENDINGS SUBTEST

Item analyses indicated that performance on all test items was satisfactory. As part of the analyses, items were grouped by inflected ending category and a mean score for each category was calculated. Items within a category were scrutinized for discrepant performance. The only notable differences in performance were on those items within a category where "none of these" was the correct response. As a result of the analysis of items by category, however, a decision was made to increase the number of items assessing the "y" ending from one to two items, and to reduce the number of items in which the correct response was a root word (from 11 to 10).

In addition, minor word changes were made in item stems for 3 of the 39 items as a result of second grade students' difficulty in reading certain words during the actual administration of the subtest.

CONTRACTIONS & POSSESSIVES SUBTEST

In order to evaluate test items in this subtest, the items were divided into two groups: contractions and possessives. Scores within each group were examined, and items on which performance was significantly lower than the mean score for the item group were identified. As a result of this procedure, two contraction items and two possessive items were revised.

Items with a correct response of "none of these" were analyzed separately to see if performance was generally poorer on these items. As in the Inflected Endings Subtest, performance by contraction category was generally poorer on these items. The correct response choice for two of the items was changed to "none of these," and two of the items where "none of these" was formerly the correct response were altered so that the actual correct response was presented. In this way, performance could be compared when "none of these" was the correct response to when it was not the correct response in the spring 1980 administration of the test. Also, the correct response choice for one possessive item was changed to "none of these." Formerly, there were no possessives items with a correct response of "none of these."

A final change affected the ordering of the items within the subtest. Those items assessing the possessive apostrophe were dispersed

throughout the test, rather than grouped together at the beginning of the test as they had been in the previous version.

AFFIXES SUBTEST

Mean scores on items within each affix category were examined, and scores on items within a single category which varied considerably from scores on other items within that same category were identified. As part of this process, each response choice was reviewed for its likelihood of being selected. For three items, an incorrect response choice appeared to be easily confused with the correct response. In these three items, a new response choice replaced the confusing foil. For all test items, familiarity level of each correct response was checked in The Living Word Vocabulary for grade level appropriateness. If there was low performance on an item, and if the grade level familiarity rating of the response word was too high, the item was revised. As a result of this procedure, 8 of the 53 target words were modified.

Another issue that was investigated was whether the inclusion of a synonym for the root word within the item stem affected children's performance. A comparison of items with and without synonyms in the stems revealed no discrepancies in performance between the two groups.

Phonics Subtests

CONSONANTS SUBTEST

Minor revisions were made in approximately 16% of the items in the preparation of the final Consonants Subtest: Three of the 59 items were

changed in the Initial Consonants section, and 11 of the 31 items were revised in the Final Consonants section. A practice item was eliminated from each section, which reduced the number of practice items to three for Initial Consonants and two for Final Consonants.

All revisions were minor; there were no items in which the target synthetic word and all four response choices were completely altered. Most of the revisions were made for one of the following three reasons:

1. More than one response choice within an item had the same medial vowel sound. New words were substituted which had the same consonant sounds in the position of interest (initial or final), but which had different medial vowel sounds.

2. Two of the incorrect response choices ended with the same consonant sound in the position of interest. One of those two response choices was changed to a new word ending with a different consonant sound.

3. [illegible] from previous versions of the Consonants Subtest showed that for consonant cluster and consonant digraph items, students often incorrectly selected the response choice that contained only one of the two letters of the cluster or digraph, especially when this foil choice appeared <u>before</u> the correct response choice. Therefore, response choices within some of the cluster and digraph items were reordered so that such foils and the correct response choice would appear first with equal frequency.

VOWELS SUBTEST

Minor revisions were made in approximately 13% of the items in preparation of the final Vowels Subtest: Five of the 51 items were revised in the Medial Vowels section, and two of the five items were modified in the Final Vowels section. As in the Consonants Subtest, all revisions consisted of only partial changes; no items were completely altered. Revisions on medial position items were made for one of the following two reasons:

1. Results from previous versions of the Vowels Subtest indicated that when students made errors, they tended to select foils containing another frequent sound of the target vowel (e.g., in an item assessing the sound of long a, students who made errors most often selected the response choice containing the sound of short a). Students were especially likely to make an error when this foil appeared before the correct response choice. Therefore, response choices within some items were reordered so that such foils and the correct response choice would appear first with equal frequency.

2. Sometimes a response choice that was intended to be acoustically close, was actually both acoustically close and visually close. Whenever possible, the response choice was changed so that it would be acoustically close only.

The two revisions made on items assessing vowel sounds in final position involved altering the target synthetic words. The target synthetic words were changed so that the initial consonant sound in the synthetic word would be different from the initial consonant sounds in any of the response choices.

CONCLUSIONS

The primary purpose of this study was to obtain item analysis information for use in the preparation of the subtests for the final administration of the Word Identification Test battery, to be conducted in spring 1980. While only minor revisions were necessary, these revisions should improve the diagnostic precision of the tests and eliminate concerns regarding correct response position and frequency of occurrence of specific distractor categories.

On the whole, the revisions described above will result in a Word Identification Test battery which is short and easy to administer, but which has sufficient scope so as to enable educators to make accurate diagnostic decisions about the apportionment of time for instruction of the various word identification skills. The revised Structure Subtests require students to make use of context in order to establish the meanings of frequently occurring elements in meaningful text. The format of the Structure Subtests is particularly suitable for teaching and assessing word identification skills in an integrated manner. The revised Phonics Subtests should enable teachers to identify specific areas of strength and weakness in children's knowledge of letter sound correspondences.

Another purpose of the present study was to evaluate the grade level appropriateness of each subtest. Some of the revisions made in Structure Subtest items eliminated words in the item stems which had caused children difficulty. Results indicated that (a) the Phonics subtests could be

administered to younger children (first grade), and (b) the Structure Subtests could be administered to older children (fourth and fifth graders for Affixes, and fourth graders for Possessives & Contractions).

In spring 1980, the final Word Identification Test battery and the Reading Subtest of the Metropolitan Achievement Tests were administered to approximately 100 children in each of grades one through five. The data will be used to examine correlations between word identification skills, as measured by the various word identification subtests, and global comprehension ability, as measured by a standardized test of reading comprehension. In addition, levels of skills mastery for each of the five subskills assessed in the battery will be established. The results of this study will be pr ited in a final report to be available in fall 1980. This final report will mark the conclusion of the research conducted over the past three and a half years by the Project on the Assessment and Analysis of Word Identification Skills in Reading (SIGH).

REFERENCE NOTES

1. Venezky, R. L. A computer program for deriving spelling to sound correlations. Master's Thesis, Cornell University, 1962. Published in part in H. Levin et al., A basic research program on reading. Ithaca, NY: Cornell University, 1963.

REFERENCES

Bouma, H. Visual recognition of isolated lower-case letters. Vision Research, 1971, 11, 459-474.

Carroll, J. B., Davies, P., & Richman, B. The American heritage word frequency book. New York: Houghton-Mifflin, 1971.

Clymer, T., et al. Reading 720, Rainbow Edition. Lexington, Mass.: Ginn & Company, 1979.

Dale, E., & O'Rourke, J. The living word vocabulary. Elgin, Ill.: Dome Inc., 1976.

Durost, W. N., Bixler, H. H., Wrightstone, J. W., Prescott, G. A., & Balow, I. H. Metropolitan achievement tests, Form F. New York: Harcourt Brace Jovanovich, Inc., 1970.

Durr, W. K., et al. The Houghton-Mifflin reading series. Boston: Houghton-Mifflin Co., 1974-1978.

Farr, R. C., Prescott, G. A., Balow, I. H., & Hogan, T. P. Metropolitan achievement tests, Form F. New York: Harcourt Brace Jovanovich, Inc., 1978.

Johnson, D. D., Pittelman, S. D., Schwenker, J., Shriberg, L. K., & Morgan-Janty, C. Relationships between word identification skills and reading comprehension of elementary school children (Technical Report No. 490). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1978.

Johnson, D. D., Pittelman, S. D., Schwenker, J., & Shriberg, L. K. The assessment of structural analysis skills (Technical Report No. 526). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1979.

Johnson, D. D., Shriberg, L. K., Pittelman, S. D., & Schwenker, J. A new approach to the assessment of phonics skills (Technical Report No. 532). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1979.

Johnson, M. S., et al. The read system (teacher's annotated e .). New York: American Book Company, 1968-1972.

Kenyon, J. S., & Knott, T. A. A pronouncing dictionary of American English. Springfield, MA: G. & C. Merriam Co., 1953.

Miller, G. A., & Nicely, P. E. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 1955, 27, 338-352.

Otto, W., et al. Wisconsin design for reading skill development. Developed at the Wisconsin Research and Development Center for Cognitive Learning. Minneapolis, Minn.: National Computer Systems, 1972-1975.

Peterson, G. E., & Barney, H. L. Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 1952, 24, 175-184.

Witty, P. A., Bebell, M. H., & Freeland, A. M. Reading caravan (Rev. ed.). Boston: Heath and Company, 1968.

Center Planning and Policy Committee

Richard A. Rossmiller
Center Director

Dale D. Johnson
Area Chairperson
Studies in Language, Reading and Communication

Marvin J. Fruth
Area Chairperson
Studies in Implementation of Individualized Schooling

Penelope L. Peterson
Area Chairperson
Studies of Instructional Programming for the Individual Student

James M. Lipham
Area Chairperson
Studies of Administration and Organization for Instruction

Thomas A. Romberg
Area Chairperson
Studies in Mathematics and Evaluation of Practices in Individualized Schooling

Associated Faculty

Vernon L. Allen
Professor
Psychology

B. Dean Bowles
Professor
Educational Administration

Thomas P. Carpenter
Associate Professor
Curriculum and Instruction

W. Patrick Dickson
Assistant Professor
Child and Family Studies

Lloyd E. Frohreich
Associate Professor
Educational Administration

Marvin J. Fruth
Professor
Educational Administration

Dale D. Johnson
Professor
Curriculum and Instruction

Herbert J. Klausmeier
V.A.C. Henmon Professor
Educational Psychology

Joel R. Levin
Professor
Educational Psychology

James M. Lipham
Professor
Educational Administration

Dominic W. Massaro
Professor
Psychology

Donald M. McIsaac
Professor
Educational Administration

Wayne Otto
Professor
Curriculum and Instruction

Penelope L. Peterson
Assistant Professor
Educational Psychology

Thomas S. Popkewitz
Professor
Curriculum and Instruction

Gary G. Price
Assistant Professor
Curriculum and Instruction

W. Charles Read
Professor
English and Linguistics

Thomas A. Romberg
Professor
Curriculum and Instruction

Richard A. Rossmiller
Professor
Educational Administration

Peter A. Schreiber
Associate Professor
English and Linguistics

G. Robert Tabachnick
Professor
Curriculum and Instruction

Gary G. Wehlage
Professor
Curriculum and Instruction

Louise Cherry Wilkinson
Associate Professor
Educational Psychology

Steven R. Yussen
Professor
Educational Psychology

11/79