INFORMATION TO USERS

This dissertation was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”.' If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value; however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106
A Xerox Education Company
MEASURING READING COMPREHENSION

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Joseph Clement Kretschmer, Jr., B.S., M.A.

* * * * *

The Ohio State University
1972

Approved by

[Signature]
Adviser

Department of Education
ACKNOWLEDGEMENTS

The author would like to express his gratitude to Drs. Charles B. Hulsenman, Jr., Victor Rentel, Robert Emans, Robert Bargar and John Kennedy for their help and encouragement in the preparation of this study. Also, gratitude is expressed to the three principals who gave permission for this research to be performed in their schools: Mr. William Garrison, Prairie Lincoln Elementary School, Southwestern City Schools, Mr. Edward Zoretic, Norton Middle School, Southwestern City Schools, and Mr. Joseph Balint, Hubbard Avenue Elementary School, Columbus City Schools. Appreciation is extended to Mr. Ted Smith, Ohio State University, for assistance with the computer programs.

JCK

1972
VITA

June 11, 1942 . . . Born - St. Louis, Missouri

1964 . . . . . . B.S., University of Missouri, Columbia, Missouri

1964-1970 . . . . Teacher of English and reading, David Wark Griffith Junior High School, Los Angeles City Schools

1965 . . . . . . Participant, NDEA Disadvantaged Youth Institute, Chico State College, Chico, California

1968-1970 . . . . Workshop leader, Los Angeles City School Districts, Division of Specially Funded Programs

1968-1970 . . . . Supervising Teacher, University of California at Los Angeles, Department of Teacher Education


1969-1970 . . . . Instructional Coordinator, Student Achievement Center, David Wark Griffith Junior High School, Los Angeles City School Districts

1969 . . . . . . M.A., California State College, Los Angeles, California

1970-1972 . . . . NDEA Title IV Fellow, The Ohio State University, Columbus, Ohio

PUBLICATIONS


FIELDS OF STUDY

Major Field: Reading Education


Studies in Curriculum and Teacher Education. Professors E. Alberty, J. Hough.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>II. RELATED RESEARCH</td>
<td>4</td>
</tr>
<tr>
<td>Reliability and Validity of Standardized Tests</td>
<td>4</td>
</tr>
<tr>
<td>Research Instruments</td>
<td>8</td>
</tr>
<tr>
<td>Reading as Reasoning: Thorndike</td>
<td>12</td>
</tr>
<tr>
<td>Reading as Problem-solving: Husbands and Shores</td>
<td>14</td>
</tr>
<tr>
<td>III. THEORETICAL ASSUMPTIONS</td>
<td>18</td>
</tr>
<tr>
<td>The Pilot Study</td>
<td>20</td>
</tr>
<tr>
<td>Paragraph Analysis</td>
<td>22</td>
</tr>
<tr>
<td>Primary addition of classes</td>
<td>22</td>
</tr>
<tr>
<td>Secondary addition of classes</td>
<td>22</td>
</tr>
<tr>
<td>Bi-univocal multiplication of classes</td>
<td>23</td>
</tr>
<tr>
<td>Co-univocal multiplication of classes</td>
<td>23</td>
</tr>
<tr>
<td>Addition of asymmetrical relations</td>
<td>23</td>
</tr>
<tr>
<td>Addition of symmetrical relations</td>
<td>23</td>
</tr>
<tr>
<td>Bi-univocal multiplication of relations</td>
<td>24</td>
</tr>
<tr>
<td>Co-univocal multiplication of relations</td>
<td>24</td>
</tr>
<tr>
<td>Comprehension criteria</td>
<td>25</td>
</tr>
<tr>
<td>Questions</td>
<td>28</td>
</tr>
<tr>
<td>General statement</td>
<td>28</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Summary</td>
<td>61</td>
</tr>
<tr>
<td>Conclusions</td>
<td>64</td>
</tr>
<tr>
<td>Reasoning in reading</td>
<td>64</td>
</tr>
<tr>
<td>Some supplementary data</td>
<td>67</td>
</tr>
<tr>
<td>Limitations of the study</td>
<td>68</td>
</tr>
<tr>
<td>Implications for future research</td>
<td>69</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>72</td>
</tr>
<tr>
<td>APPENDIXES</td>
<td>76</td>
</tr>
<tr>
<td>APPENDIX A: The Pilot Study Test</td>
<td>76</td>
</tr>
<tr>
<td>APPENDIX B: Pilot Study Statistics</td>
<td>78</td>
</tr>
<tr>
<td>APPENDIX C: The Experimental Test—Form A</td>
<td>79</td>
</tr>
<tr>
<td>APPENDIX D: The Experimental Test—Form B</td>
<td>84</td>
</tr>
<tr>
<td>APPENDIX E: The Experimental Test—Form C</td>
<td>89</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mean Standardized Test Scores of Subjects</td>
<td>33</td>
</tr>
<tr>
<td>2. Means and Standard Deviations for all Subjects on Standardized and Experimental Tests</td>
<td>49</td>
</tr>
<tr>
<td>3. Correlations between Standardized and Experimental Test Scores for Third Grade Subjects</td>
<td>51</td>
</tr>
<tr>
<td>4. Correlations between Standardized and Experimental Test Scores for Sixth Grade Subjects</td>
<td>52</td>
</tr>
<tr>
<td>5. Mean Scores for Question Categories</td>
<td>58</td>
</tr>
<tr>
<td>6. Analysis of Variance of Experimental Test Categories for Both Grades</td>
<td>58</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

FIGURE PAGE

1. Digraph of relationships for examples in paragraphs 1 and 2 .................................................. 25
2. Digraph of relations for example in paragraph 7 .................................................................................. 25
3. Digraphs showing cases where criterion (a) or (b) are used to assess difficulty of comprehension .......................................................................................................................... 26
4. Additive classification paragraph and accompanying digraph .......................................................... 35
5. Multiplicative classification paragraph from Thorndike's (1917) study and accompanying digraph ......................................................................................................................... 36
6. Additive series paragraph and accompanying digraph ........................................................................ 37
7. Multiplicative series paragraph and accompanying digraph ................................................................ 37
8. An easy and a hard item as judged by criterion (a) from Paragraph CA of the experimental test. The Asterisk* denotes correct responses ......................................................................................... 39
9. Sample intension and extension-based items from Paragraph CA ................................................................ 40
10. Matrix of item descriptors for the experimental test ............................................................................ 41
11. Item descriptors for items in classificatory and serial categories ....................................................... 45
CHAPTER I

THE PROBLEM

In a preliminary statement of the general plan of action of the National Right to Read Program (The Reading Teacher, 1971), the major goal of the program was "...to substantially increase functional literacy...[p. 215]." Functional literacy was defined as "...the ability to read to the end that the individual is able to function productively as an adult and thereby increase the benefits to be derived from this society[p. 215]."

For reading to enable one "to function productively as an adult," it must fulfill for this individual a variety of purposes. These purposes for reading exist within the context of a society that is rapidly changing, and thus the purposes change just as rapidly. For many people functional literacy connotes merely being able to call off the words of a text, perhaps gaining the literal meaning of the passage. But the society in which adults function today demands much more from reading. A reader must wade through a morass of printed information that can never be digested in the slow, methodical manner devoted to novels in the Victorian period. The reader must be prepared to
change his methods of reading as he encounters materials that continue to become ever more specialized in their purposes.

As reading materials continue to diversify, the list of "reading comprehension skills" grows into huge taxonomies which become so cumbersome that a teacher is frustrated in her search for classroom activities to foster reading comprehension. The only adequate solution is to facilitate the growth of a reading "intelligence" or the keenly developed group of thinking abilities so modified that they react to data in print as smoothly as they do to concrete experiences. This mental set would be extremely flexible, and thus applicable to the vast and rapid changes that are expected in the future. Alvin Toffler, author of *Future Shock* (1970), quotes psychologist Herbert Gerjuoy of the Human Resources Research Organization:

> The new education must teach the individual how to classify and reclassify information, how to evaluate its veracity, how to change categories when necessary, how to move from the concrete to the abstract and back, how to look at problems from a new direction—how to teach himself. Tomorrow's illiterate will not be the man who can't read; he will be the man who has not learned how to learn [p. 414].

One of the reasons for the plethora of skills lists for reading comprehension is the lack of knowledge about the basic process itself. Simons (1971) remarks:

> That comprehension research is at this 'early
fact gathering stage with the concomitant inability to separate relevant from irrelevant facts is evidenced by the proliferation of comprehension skills and tests to measure them and the inability to distinguish comprehension from other psychological processes [p. 355].

This lack of knowledge about the basic process involved in reading comprehension not only thwarts attempts to provide curricular experiences for students, but also blocks accurate measurement of understanding in reading. Roger Farr, chairman of the Test Evaluation Committee for the International Reading Association, remarked in a 1969 monograph that "...no one seems to know whether sub-skills of reading can be measured, and that there is a lack of measures for assessing more complex reading behaviors...[p. 71]."

Research in reading comprehension, as any other research, depends on accurate measures. If theorists and measurement specialists wait upon each other to provide the essential insight into the nature of reading comprehension, then research is clearly doomed to remain at a standstill.

Research in the measurement of reading comprehension is intimately connected to research into the basic process itself. This study attempts to deal with the reading comprehension process by constructing an experimental instrument to measure thinking within the context of reading.
CHAPTER II

RELATED RESEARCH

Research in the measurement of reading comprehension can be divided into two categories: reliability and validity studies, and attempts to construct new instruments. In addition to these studies, two articles that are particularly relevant to this study are discussed.

Reliability and Validity of Standardized Tests

Standardized reading comprehension tests are generally subtests of other reading tests. A typical reading test is composed of two subtests: vocabulary and comprehension. There may also be a speed subtest. Unlike other subtests, comprehension subtests many times are given a variety of labels such as "paragraph reading," "paragraph comprehension," "paragraph meaning," "paragraph understanding," "critical reading," etc. Although the usual format for a comprehension test consists of paragraphs followed by multiple-choice questions, there is considerable variety. Comprehension in the Gates-MacGinitie Test is tested by having the examinee choose words to insert in several blanks in the paragraph, similar to the cloze...
procedure. The Monroe-Sherman Test presents a single question above the paragraph to be read, with the answer choices following the paragraph. In short, comprehension tests are by no means standardized with respect to form, and, as Livingston (1972) reported, a pupil's score can vary one to three grade levels from test to test.

Comprehension subtest reliability has been criticized. Farr (1969) noted that many comprehension subtests are too short to have adequate reliability regardless of the method used to obtain a coefficient. Most teachers could not find out which method was used anyway, as the same writer also states most test manuals fail to report subtest reliabilities. With regard to split-half reliability estimates, Thorndike and Hagen (1961) point out that they become meaningless when a test is highly speeded. As most reading tests are timed, examinees are encouraged to hurry, which may have the effect of speeding the test. Split-half reliability also shortens a test which, as was already mentioned, is generally too short to begin with.

Several studies have cast doubt on the validity of standardized reading comprehension subtests. Farr (1969) restates the generally acknowledged phenomenon that a pupil's performance on a standardized test runs about two grade levels above his performance on an individually administered reading inventory. Murphy and Davis (1949) reported a negative correlation between ability to reason in reading and academic achievement of college freshman.
It has also been shown that subjects who do not read the prerequisite passages before answering the questions of standardized comprehension subtests can correctly answer as many as 67 per cent as many items as those who do read the passages. This lack of "reading dependency" was shown by Tuinman (1970) for the STEP Reading Test, Preston (1964) for the reading subtest of the Cooperative English Test, and by Mitchell (1967) for subtests of the Gates-MacGinitie. Kretschmer (in press) surveyed 16 standardized subtests of comprehension, and found a disproportionate amount of certain subject areas such as animals and agriculture, as well as items for which many elementary pupils could have no conceptual background. Livingston's article (1972) also criticizes comprehension subtests of the Stanford for basing their selections on such topics as architectural design, respiratory functions, and the astrological origins of the names of the week. The same author condemns the Metropolitan and Iowa Reading Tests for the same reason.

In addition, Livingston continues to show how the questions on comprehension subtests make frequent use of signal words such as "always," "never," "all" or "none" which test-wary students quickly recognize. He claims 40 per cent of the Iowa Silent Reading Test (comprehension subtest) questions use these words. Using a sample paragraph from the Stanford Reading Test, he also shows
how the examinee is really tested on his ability to interpret strings of awkwardly-constructed sentences, as many test paragraphs are poorly written.

While Livingston attacks comprehension tests on their lack of content validity and face validity, Simons (1971) questions the construct validity of such tests. He reasons that since there is no clear-cut theory of reading comprehension at the present time, none of the tests purporting to measure this skill can have adequate construct validity. Instead, many test constructors use lists of "comprehension skills" that attempt to describe the behaviors involved on the part of the reader. Simons shows that these lists often include teaching procedures (e.g., underlining the key words in a paragraph) and uses for comprehension (e.g., checking author's sources) as well as psychological processes. Skills lists vary in length and content from authority to authority, and as such hardly constitute a valid model from which accurate measures can be constructed. Farr (1969, p. 64) also agrees that the most pressing research need in the measurement of reading comprehension is a clear understanding of the nature of this phenomenon, i.e., a workable theory upon which to base measurement.

In general, reading comprehension subtests of standardized test batteries have been criticized on the following points:

Lack of reliability due to (a) influence of timing
and (b) the short length of the tests, and lack of validity due to (c) the fact that these tests so often overestimate a student's performance on individual oral tests, (d) a lack of reading dependency, (e) a poor selection and distribution of content, (f) the over-use of signal words, (g) generally poor quality of writing in many paragraphs, and (h) the lack of a clear-cut theory of reading comprehension upon which to base assumptions of behavior to be measured.

Given the dubious reliability and validity of standardized comprehension subtests, it is evident why better-known reading researchers have not utilized these instruments, but constructed their own tests to fit their individual goals.

Research Instruments

In constructing their own tests, reading researchers usually chose one or more of the "comprehension skills" such as those cited by Simons (1971) above. For example, J. E. Davis (1969) recorded his subjects' abilities to distinguish fact from opinion in ten paragraphs. Koenke (1968) attempted to measure his subjects' abilities to extrapolate the main idea of a paragraph by recording their responses on a 7-point "main idea scale" with general and specific extremes.

These tests were more useful than the standardized subtests from the standpoint of analyzing reading
comprehension, for they narrowed their scope to a single aspect of reading behavior, which could then be examined closely. Unfortunately, these experiments did not isolate or estimate the particular contribution of the skill under examination to the total comprehension process.

Perhaps the most well-known researcher in the measurement of comprehension is Frederick B. Davis (1944, 1968). Davis constructed comprehension tests based on what skills he thought might be crucial to understanding text, and administered them to college freshman. He then factor analyzed the results, and concluded that comprehension was not a unitary skill, but was dependent on separate abilities, particularly knowledge of word meanings. Holmes and Singer (1964) also used this approach in building their theory of reading behavior. The problem with the factor-analytic approach is, of course, that the results of such studies are only generalizable to the type of reading behavior measured by the specific items on the tests. The only definitive statement that can be made about F. B. Davis's studies is that college freshman need to know word meanings to score high on Davis's test. No inferences for these subjects' everyday reading behavior are justified.

In 1956 Piekarz constructed 30 questions based on the skills that F. B. Davis had "identified" from his studies. In an interview situation, she asked 22 sixth-graders
these questions after they read a selection. She was able to analyze their answers to a greater degree than was possible with Davis's paper-and-pencil tests, and discovered the impact that each individual's emotional reaction to the content of the text had on their understanding of it. But she used only Davis's list of skills, and limited her subjects' readings to a single selection.

Other researcher's investigations have examined limited cognitive activities in a reading setting. Gans (1940) required her subjects to judge paragraphs as to their relative usefulness in solving a problem. Maw and Maw (1962) had pupils mark statements "foolish" or "not foolish," and Tatham (1970) had second and fourth-graders point to one of three pictures that purported to depict action in a sentence they had read. Like the studies mentioned previously in which a single comprehension skill was analyzed, they contributed but little to knowledge of the total process of comprehension.

Researchers have recently turned to the field of linguistics for answers to the comprehension problem. Simons (1971) recommends that linguistics be the basis of all serious theorizing in reading comprehension in the future. Linguistics plays a prominent part in F. Smith's (1971) theory of reading comprehension, which is supplemented with aspects of information-reduction theory.
Linguistic theories of comprehension lean heavily on an instrument known as the "cloze procedure" (Taylor, 1953), in which subjects try to guess words deleted in a random or fixed order from a reading selection. Bormuth (1966) used this procedure in his study of the relationships between linguistic variables and comprehension difficulty.

There are two problems inherent in the use of the cloze procedure for measurement of comprehension. First, only one method has been consistently used to check its validity, correlation with standardized comprehension subtests. Bormuth, for instance, used the California Reading Test as a correlate, and Ruddell (1965) used the Stanford Reading Test which Livingston (1972) objects to so strenuously. Since the reliability and validity of standardized comprehension subtests has been shown to be less than acceptable, correlations with these tests are hardly convincing evidence for cloze validity. A second objection to cloze tests is the examinee's response itself. Insertion of words randomly deleted from a text is at best only a measure of literal comprehension. The subject's task is to complete the sentence, not to analyze the reading selection for complete understanding or critical reaction.

Herein lies the crux of the problem facing researchers who take a linguistic approach to reading comprehension; linguistics simply does not provide for textual analysis.
beyond the sentence level. Semantics probes the meanings of words or morphemic units, and linguistics studies the structure of language through analysis of the sentence; but as soon as several sentences are combined to produce a piece of prose with a certain purpose, the analytical "buck" gets "passed" on to the philosophers and the psychologists. Thus basic research into the nature of reading comprehension has been done by psychologists such as Ausubel (1963), Frase (1967; 1968; 1969), Berlyne (1966) and Rothkopf (1966); and the first noteworthy investigation of reading comprehension was done by another psychologist, Edward Lee Thorndike.

Reading as Reasoning: Thorndike

Probably no single study of reading comprehension has been cited as frequently as that of E. L. Thorndike (1917). Its simplicity and impact have made it a classic, and it is still discussed today. Recently, an issue of the Reading Research Quarterly (v. VI, Summer 1971) reprinted one of Thorndike's three original reports of the study, along with commentary articles by Wayne Otto and Russell G. Stauffer.

In format, the study was strikingly simple. Thorndike wrote several short paragraphs each followed by five to seven questions over the information in each paragraph. The subjects (children in grades five through eight) answered these in their words by writing in the spaces after each cue.
question. What he received was a plethora of answers for each question, ranging from accurate to ridiculous, the great majority of them incorrect. This stunned the educationists of the day, since the questions seemed so easy to the average literate adult. Thus Thorndike was able to argue convincingly that reading was indeed reasoning, and much more complex than had been thought previously.

But the main impact of Thorndike's study, in the words of Wayne Otto (1971) was that "...many of the very questions Thorndike left unanswered in his report remain unanswered—or unanswerable in any definitive way—to this day. [p. 440]." This remark was directed in particular to the poor performance of Thorndike's subjects, for Otto found that even today, with literacy much more widespread, children's formulation of the main idea of a paragraph is relatively poor when rated on a scale developed for use in his study (Otto & Koenke, 1970).

But if children's performance on paragraph comprehension is generally poor, then why do they constantly score higher on standardized tests than on informal inventories? Again, Otto (1971) states that his children's mistakes in paragraph reading were not "...with remembering and/or organizing facts and principles but with understanding them [p. 441]." Standardized comprehension subtests ask many questions that deal with the former than the latter. While many reading experts have mouthed the phrase
"reasoning in reading," few have attempted to construct instruments to measure a reader's ability to think when printed text is the stimulus, and Thorndike's fifty-five-year-old study is still without adequate follow-up.

**Reading as Problem-solving:**
Husbands and Shores

Thirty-three years after Thorndike's study, Husbands and Shores (1950) reviewed the literature in measurement of reading to solve problems. Their comments on the state of test development in reading comprehension is still accurate today:

...the majority of research leading to the development of reading tests has been previously concerned with the measurement of level of reading ability in terms of the ability to grasp and retain facts contained within a relatively short paragraph. There is a dearth of creative research designed to construct devices for the measurement of reading as a tool for the understanding of relationships or for the solution of problems p. 453.

They continue to describe how factors such as purpose, rate, content, etc., were generally ignored by test constructors—a point echoed in Farr's 1969 monograph and Livingston's 1972 article. But by far the most salient statement is the following:

*Teachers want to know whether children can read and think about what they read. The total score would then represent neither reading nor thinking but a new measured*
ability to think when reading material provided the basic data for the thought process [p. 459, emphasis investigator's].

It is understandable why researchers have been reluctant to construct such an instrument as the one described by Husbands and Shores. One particularly vexing problem is that a circular argument develops when attempting to establish construct validity for such a test, i.e., which should come first--accurate measures or basic theory? Measures should be based on theory, but psychological theory should be based upon empirical observations that rely, in turn, upon accurate measures.

One way to break this cycle might be to use theory from a closely-related area that is based upon a substantial amount of empirical observation as a basis for constructing a test of thinking in reading. This is the method adopted by researchers such as Goodman (1970) and Ruddell (1970), who apply linguistic and psycholinguistic principles to the study of reading behavior. It is also followed by Lawrence Frase (1967; 1968; 1969) who uses S-R psychology and Hullian models to generate theories of how humans learn from prose text. Indeed, Thorndike started the trend when he interpreted reading in terms of his own connexionistic theory of learning. What is puzzling is that no one has seen fit to interpret reading in terms of any of the developmentally-oriented
psychological theories. This lack has persisted even though the term "developmental reading" has been in use for many years. Reading experts have applied various physical and emotional development principles to reading instruction, particularly those referring to eye and hand coordination, and the value of motivation and success experiences in reading. But developmental learning theories have not been tapped for application to reading, especially in the case of reading comprehension, where they would be most appropriate.

Could the poor performance of Thorndike's and Otto's subjects in paragraph reading be the result of incomplete mental development, rather than faulty reading behavior? Would this explain the fact that accurate paragraph interpretation by these children was a rarity? Perhaps the need is to redefine reading comprehension in light of what is known about children's thinking, and design evaluation instruments to conform to their modes of thought, rather than to those of adults, as is presently done. Simply shortening sentences and using "easy" words may not facilitate a task that is beyond the mental development of the reader.

One developmental theory that is based solely upon thousands of empirical observations of children at all ages is that of the Swiss psychologist, Jean Piaget. His theory is most fitting for application to reading since, as Flavell (1963) notes, "a persistent and
overriding interest in the area of intelligence is a salient feature distinguishing Piaget's work from that of most child psychologists [p. 16]." Another distinguishing feature is that his is probably the only epistemological theory that derives its empirical basis from children's actions, rather than those of adults or college sophomores, as is common to most American research in learning. This makes Piaget's theory most relevant to research in reading, since the bulk of reading instruction in the United States takes place in elementary schools. Thus it is Piaget's models of child logic which will provide the basis for construction of an experimental test of reading comprehension for use in this research.
CHAPTER III

THEORETICAL ASSUMPTIONS

B. Inhelder and J. Piaget (1964) have observed classificatory and serial activities in children ranging in age from three to sixteen years. They concluded that during the ages seven to eleven, "concrete" forms of logic develop in three distinct stages irregardless of material the subjects used as data.* Their subjects used varied materials such as geometric shapes, dolls and toys, pictures of animals and flowers, etc.

This research proposes a model of reading comprehension based upon cognitive activities observed by Piaget and others in children from seven to eleven years of age. This is the period of concrete operations, wherein children's logical powers are concentrated in eight mental operations, four classificatory and four serial. But the unique nature of reading activity prevents a simple transposition of Piagetian tasks to a

*This is not to say that the nature of the problems data had no effect on the subjects' abilities to classify and seriate. When material became more abstract, e.g., pictures of animals, development was delayed, but the order of stage development remained the same.
reading format. For example, it may be argued that the concrete operations (so called because they are based on activities with concrete props) are no longer concrete, but have been reduced to verbal problems. Piaget (1967) himself states that:

If a child at this level is asked to reason about simple hypotheses, presented verbally, he immediately loses ground and falls back on the prelogical intuition of the preschool child. For example, all children of nine and ten can arrange colors into a series even better than they can arrange sizes, yet they are completely unable to answer the following question, even when it is put into writing: 'Edith has darker hair than Lily. Edith's hair is lighter than Susan's. Which of the three has the darkest hair?' In general, they reply that since Edith and Lily are dark-haired and Edith and Susan are light-haired, Lily is the darkest, Susan the lightest, and Edith in between. On the verbal plane, they succeed in producing only a series of uncoordinated pairs, as children of five or six do when they attempt to seriate a set of size-graded objects...If the children were able to manipulate objects, they would be able to reason without difficulty, whereas apparently the same reasoning on the plane of language and verbal statements actually constitutes other reasoning that is much more difficult because it is linked to pure hypotheses without effective reality [p. 62, emphasis investigator's].

This problem, from some old tests of Burt's (1913), has been quoted extensively to support Piaget's contention that concrete logic is possible only with support of concrete props. Such a statement, if true, would cast doubt on the efficacy of the entire notion of basing a reading comprehension test on Piagetian operations, since reading deprives subjects of concrete props. Moreover,
the statement that this inability occurs even when the question is in written form (i.e., requiring reading) is even more damaging.

This statement, together with Thorndike's (1917) and Otto's (1970) observations of poor performance by children in paragraph comprehension, led the investigator to do a pilot study to test the feasibility of children succeeding on a reading test involving concrete operations.

The Pilot Study

Forty-five fourth-graders (20 boys, 25 girls) in two intact classes at Hubbard Avenue Elementary School, Columbus City Schools, were selected as subjects for the study. These children were nine and ten-year-olds, the same ages as those children whom Piaget observed were unable to answer the hair-color problem.

A twenty-item test was constructed, with four paragraphs that represented Piagetian tasks at the concrete operational level. Three of these were serial, and one was classificatory. The entire results will not be reported here, since the instrument in the major study yielded more reliable data. The complete test and performance statistics for the pilot study can be found in appendixes A and B of this paper.

Immediately relevant to the point in question, i.e., Piaget's contention that children cannot apply concrete logic in the absence of concrete props, is the performance
of the children on the first two items of the pilot study test. These items were open-ended questions over the following short paragraph:

    Jim runs faster than Ted.
    Dan runs slower than Ted.

This statement is logically identical to the problem about hair color, i.e., it requires subjects to establish an asymmetrical relationship among three objects (people, in this case). The subjects were asked the following questions over this paragraph, and wrote their answers in their own words under each question:

1. Who runs the fastest?
2. Who runs the slowest?

Contrary to what Piaget observed in the hair-color problem, 98 per cent of the 45 subjects wrote a correct answer to the first question (Jim), and 93 per cent gave the correct response for item number 2 (Dan). In addition, the subjects proved capable of reasoning in even more complex paragraphs that required longer serial chains. The mean per cent of correct responses for ten items over three serial paragraphs was 73.6, a rather respectable figure, which becomes even more meaningful when one considers that the subjects, according to their principal and teachers, read an average of two grade levels below national norms.

On the basis of this pilot study, the investigator
decided that the construction of a reading comprehension test based upon concrete operations was indeed feasible, and proceeded to construct a method by which these operations could be delineated in prose text, and set against criteria of difficulty. This method was called \textit{Paragraph Analysis}.

\textbf{Paragraph Analysis}

Paragraph Analysis (PA) is a procedure by which a test constructor can classify the information in a written paragraph as requiring either classificatory or serial activity by the young reader. He then constructs questions to test for evidence of these behaviors.

According to Flavell (1963) Piaget proposes eight types of concrete operations performed by the child of seven to eleven years of age:

1. \textbf{Primary addition of classes} is simply the construction of a class hierarchy. A sentence such as "John has a tall brother and a short brother" establishes a class of people that includes John and his brothers, and two subclasses of tall and short members within the subclass "brothers of John."

2. \textbf{Secondary addition of classes} is the inclusion of the complements of each class and subclass in the established hierarchy. Thus the existence of a class "John's brothers" necessitates existence of its complement, or secondary class,
"those who are not John's brothers" which includes John.

3. **Bi-univocal multiplication of classes** occurs when a class hierarchy is established classifying identical members according to two or more attributes. John's brothers might be classified according to hair color, weight, disposition, etc., as well as height.

4. **Co-univocal multiplication of classes** is the situation in paragraph three, with the difference that not all multiplications are possible. For example, subclasses established on the properties of "being John" and "not being John" might also be classified on the properties of "being John's brother" and "not being John's brother." Obviously, a subclass cannot exist that has simultaneous properties of "being John" and "being John's brother."

5. **Addition of asymmetrical relations** sets up an ordered difference scale with objects or classes occupying various points. An example would be three articles of jewelry on a monetary value scale where A (a necklace) costs more than B (a ring) which costs more than C (a bracelet). It is necessary to conclude that A costs more than C.

6. **Addition of symmetrical relations** is analogous to the example in paragraph 5, but without differences, i.e., every member is equivalent to each other with respect to the scaled value. Thus A (a book) costs the same as B (a pen) which costs the same as C (a wallet). It is
necessary to conclude that A costs the same as C.

7. **Bi-univocal multiplication of relations** compares a series of elements on two or more ordered difference scales, just as bi-univocal multiplication of classes classifies a group of elements on two or more attributes. If A (the necklace), B (the ring) and C (the bracelet) were compared on a scale of weight as well as cost, so that A costs more and weighs more than B which costs more and weighs more than C, this would be a bi-univocal multiplication of relations, or bi-univocal seriation.

8. **Co-univocal multiplication of relations** compares elements with respect to both symmetrical and asymmetrical relations. Thus if an element were added to the articles of jewelry, e.g., D, so that D costs the same but weighs more than B, or some other combination of equivalence and difference, co-univocal seriation would result.

To obtain economy and clarity, PA uses digraphs (Barary, Norman & Cartwright, 1965) to build models of classes or relations in the paragraph. In digraphs of classes, points with upper-case letters represent classes and subclasses, and the arrowed lines represent inclusion relationships. Thus B → A is read "class B is included in class A." Figure 1 shows a digraph for the example classes in paragraphs 1 and 2.

Digraphs for relations (e.g., paragraphs 5 to 8) need to be slightly modified. Points and upper-case letters represent the elements compared, and the arrowed lines
Fig. 1. Digraph of relationships for examples in paragraphs 1 and 2.

represent the relationships. In cases where two or more relationships are involved, e.g., paragraphs 7 and 8, the lines must be labelled as well as the points. This is done with lower-case letters to indicate the relations. Figure 2 shows a digraph of relations in paragraph 7:

\[ N \text{(necklace)} \rightarrow R \text{(ring)} \rightarrow B \text{(bracelet)} \]

\[ a_1 \rightarrow b_1 \rightarrow B \]
\[ a_2 \rightarrow b_2 \]

\[ a = \text{costs more than} \]
\[ b = \text{weighs more than} \]

Fig. 2. Digraph of relations for example in paragraph 7.

Comprehension criteria. PA uses two criteria in estimating the difficulty of comprehending written prose. All information in a given paragraph is analyzed for both classificatory and serial relationships. Digraphs are then
drawn to serve as reference models. Two criteria are then applied to the models as an estimate of difficulty. These criteria are: (a) number of logical steps (arrowed lines) in cases where addition of classes or relations are involved, and (b) number of properties or ordered difference scales referred to in identifying an element in cases where multiplication of classes or relations is involved. Figure 3 shows how these criteria are depicted in digraph form:

Fig. 3. Digraphs showing cases where criterion (a) or (b) are used to assess difficulty of comprehension.

Figures 3a and 3c are similar, and thus both use criterion (a), the number of lines between elements in the digraphs (which represent logical steps). Using the symbols \(\supset\) and \(\subset\) as "includes" and "is included in" and
> and < as "greater than" and "less than," we can state assertions. In Figure 3a, D ⊃ C and C ⊃ B are predicted to be easier to comprehend than D ⊃ B or D ⊃ A, because there are more logical steps required to conclude the latter two assertions in the model. Frase (1969) used digraphs to model paragraph content in his study, and called this logical number of steps "structural distance," and the classes "text points." He found that the greater the structural distance between two text points, the less often it was recalled.

Likewise, in Figure 3c, D > C and C > B should be less difficult than D > B and D > A. For multiplication (Figures 3b and 3d) the criterion is number of properties or ordered difference scales. Thus, in Figure 3b, it should be easier to understand B ⊃ A₁ than B ⊃ A₁A₂, which in turn is less difficult than C ⊃ A₁A₂, which combines logical distance with number of properties (although this combination of the two criteria will not be examined in this study). In Figure 3d, it should be easier to comprehend a statement like D →₃ C or D →₃,₂₂ → B than statements like D →₂₁ →₃₁ B or D →₃₂,₂₂ →₁₃₁₂₁ B, G.

These predicted differences in understanding were not found by Inhelder and Piaget in their 1964 study. But their subjects worked with concrete props. Reading is a more abstract task, although it is not known exactly what makes it more abstract. Finding out whether or not these predicted differences in comprehension difficulty hold true
will be the first step in the examination of how reading modifies the type of cognitive activities that are usually associated with direct experiences.

Questions

General statement. It seems logical to conclude that as a task becomes more abstract, it also becomes more difficult. Piaget (1964, p. 247) has stated that classes are innately more abstract than series, as the former cannot be directly perceived while the latter represents a "good form" perceptually. He and his co-workers found no differences in the development of children's ability to classify and seriate, but in reading classification may prove to be more difficult.

Differential wordings of the questions Piaget and his associates asked their subjects had no effect upon their findings. But in a reading task like a comprehension test, the subjects must read and comprehend the questions as well as the paragraphs. Aside from keeping the readability of the questions, i.e., difficulty of words and sentence length, commensurate with that of the paragraph, there is another factor in question construction that could influence comprehension. Classes and series can be conceptualized in terms of intension or extension. Intension refers to the property which defines a class or subclass, e.g., height, hair color, etc. in the
case of John and his brothers. Extension refers to the members of the class or subclass, e.g., Fred, Will, John, etc. Relations can also be analyzed in this manner, with the higher or lower cost in the jewelry example being the intension, and the individual articles (necklace, ring and bracelet) being the extension. If the question stem (in a multiple-choice item) is constructed in terms of the intension of a class or series, would the subjects not be forced to think more abstractly than if the question listed the extension of the class or series?

There is also a question as to whether paragraphs that convey only additive class or relation hierarchies are not, as a whole, easier to comprehend than paragraphs which set up multiple hierarchies of classes or relations.

Since Piaget's theory is developmental in nature, the older subjects should perform significantly better in all areas.

Finally, if the experimental instrument truly reflects ability to apply thinking skills to reading, relationships between standardized intelligence and reading tests should be assessed. These relationships should be relatively low since standardized instruments have not been designed to evaluate a Piagetian definition of thinking in reading.

There are thus eight specific questions:

Question 1. Will there be significant relationships between experimental test performance and IQ as measured
by traditional instruments?

**Question 2.** Will there be significant relationships between experimental test performance and reading vocabulary or reading comprehension scores on traditional instruments?

**Question 3.** Will there be significant differences between experimental test performance of third-graders and sixth-graders?

**Question 4.** Will there be significant differences between subjects' performance on classification items and seriation items of the experimental test?

**Question 5.** Will there be significant differences between subjects' performance on items following additive paragraphs and items following multiplicative paragraphs of the experimental test?

**Question 6.** Will there be significant differences between subjects' performance on experimental test items judged easy by criterion (a) or (b), and on experimental test items judged hard by criterion (a) or (b)?

**Question 7.** Will there be significant differences between subjects' performance on experimental test items whose stems are constructed in terms of classificatory or serial intension, and those whose stems are constructed in terms of classificatory or serial extension?

**Question 8.** Will there be significant interaction among subjects' performance on any of the categories of items of the experimental test?
CHAPTER IV

PROCEDURE

General Procedure

The overall strategy of this research is to construct a reading comprehension test based upon the eight concrete operations that, according to Inhelder and Piaget (1964) and Piaget (1967), form the basis for child thought between the chronological ages of seven and eleven. This test would then be administered to two groups of subjects, one older and one younger, within the seven-to-eleven age range to assess (1) the ability of these subjects to apply concrete operations to data in a reading format, and (2) any differences with regard to the type of paragraph (classificatory or serial, additive or multiplicative) or the type of question (easy or hard, intension-based or extension-based). Relationships between the experimental test and subjects' IQ, reading vocabulary and reading comprehension scores on standardized tests would also be investigated.

Selection of Subjects
In selecting subjects, four factors needed to be considered:

1) Two groups of subjects are necessary, far enough apart in age to show developmental differences, yet old enough to have few word recognition problems when reading the experimental test.

2) Subjects should be normally distributed with respect to mental abilities.

3) Subjects should be relatively homogeneous with respect to socioeconomic status, cultural background and dialectical differences.

4) Groups should be large enough to allow statistical tests to be performed on the test data, and to represent a fair proportion of the population to which the results will be generalized.

Sixty third-graders (two intact classes) from Prairie Lincoln Elementary School in Southwestern City School District in Franklin County, Ohio, and 60 sixth-graders from Norton Middle School in Southwestern City School District in Franklin County, Ohio were chosen as subjects. Only sixth-graders who have previously attended Prairie Lincoln Elementary will participate. There are approximately the same number of girls as boys in each group. Table 1 shows the mean scores on the Lorge-Thorndike Intelligence Test for both groups, and the reading vocabulary and comprehension subtests of the Iowa Silent
Reading Test for the third-graders, and the Paragraph Understanding subtest of the Monroe-Sherman Reading Test for the sixth-graders (reading vocabulary scores were not available for this group).

### TABLE 1

<table>
<thead>
<tr>
<th>Grade</th>
<th>IQ</th>
<th>Reading</th>
<th>Voc.</th>
<th>Comp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third</td>
<td>102.2</td>
<td>3.42</td>
<td>6.71</td>
<td>3.71</td>
</tr>
<tr>
<td>Sixth</td>
<td>105.2</td>
<td>-</td>
<td>6.65</td>
<td></td>
</tr>
</tbody>
</table>

**Test Construction**

A thirty-two item multiple-choice test was constructed, based on the eight concrete operations previously described. It consisted of four paragraphs with eight questions over each paragraph.

The paragraphs. To serve as data for the classificatory and serial tasks of the test, four short paragraphs were used. Three were written by the investigator, and one was selected from the well-known experiment by Thorndike (1917) discussed earlier in this paper. Each paragraph established one of the following operations:

1. an additive classification
2. a multiplicative classification
(3) an additive series
(4) a multiplicative series

To render each paragraph as homogeneous as possible with regard to content and readability, each paragraph (a) was between 43 and 51 words in length; (b) discussed relationships among a group of boys and girls; (c) contained no words (except names) that were not on the Dale List of 3000 Familiar Words (Dale & Chall, 1948); (d) had an average of 5.7 sentences, with an average sentence length of 8.15 words. The combination of short sentences and absence of "hard" words, i.e., those not on the Dale List, placed the readability of the test paragraphs below the lower limit of the Dale-Chall Formula (fourth-grade level). Since the average reading vocabulary and comprehension scores for the third-graders were 3.43 and 3.71 respectively, they experienced few problems in reading the paragraphs. In addition, each child had the option of asking the examiners to pronounce any word that they did not immediately recognize, and there was no time limit for completion (see Administration Procedures).

Addition of primary classes was established in a 51-word paragraph concerning five children. Figure 4 shows the text of the paragraph and the digraph of class inclusion. Actual test forms used for this study can be found in Appendixes C, D and E of this paper.

Multiplicative class relationships are shown in the
paragraph used in Thorndike's 1917 study. It is 43 words long, and tells of family relationships among four siblings. Figure 5 shows the text of this paragraph and a digraph of its contents.

Betty, Brad, Dan, Greg and Karen go to Elm St. School. Greg is in the fifth grade, and the others are in fourth. All the fourth-graders like ice cream except Karen. Betty does not have a pet. The other fourth-graders who like ice cream do have pets. Dan has a dog.

Fig. 4. Additive classification paragraph and accompanying digraph.

Since certain items of information in this paragraph are unknown, e.g., whether John is tall. Classes such as D and D', (tall and short children) which would ordinarily constitute A rather than B, i.e., John (B) has to be either tall or short, are instead subsumed under a subclass (B'). Situations such as these distinguish reading comprehension from classification with simple concrete props.

Like the additive classification paragraph, the 44-
John had two brothers who were both tall. Their names were Will and Fred. John's sister, who was short, was named Mary. John liked Fred better than either of the others. All these children except Will had red hair. He had brown hair.

\[
\begin{array}{c}
A \\
E' \\
E \\
B \\
B' \\
C(D) \\
C'(D') \\
F' \\
F
\end{array}
\]

A = children in John's family  
B = John  
B' = children who are not John  
C = John's brothers  
C' = John's sisters  
D = tall children  
D' = short children  
E = red-haired children  
E' = brown-haired children  
F = Fred  
F' = Will

Fig. 5. Multiplicative classification paragraph from Thorndike's (1917) study and accompanying digraph.

word additive series paragraph concerns five children. The common relationship is distance from a first-person character. Figure 6 shows the text of this paragraph, and the digraph of relationships. The lines of the digraph have been labelled, since there is more than one relationship possible.

Multiple relationships are illustrated in the multiplicative series paragraph, shown with its digraph in Figure 7. Forty-eight words long, it also concerns a
I have five good friends. Bert lives right next door to me. Carol lives farther away than Bert. Jill lives even farther away than Carol. Dana lives farther away than Carol, but closer than Scott. I have to take a bus to his house.

Fig. 6. Additive series paragraph and accompanying digraph.

There are nine boys on the Bears baseball team. Mark, Bill and Ed are all taller than Henry or Dave or Bob. Jack, Ray and Al are the Shortest. Henry, Jack and Mark run faster than Dave, Ray or Bill. Bob, Ed and Al are the slowest runners.

Fig. 7. Multiplicative series paragraph and accompanying digraph.
group of children, in this instance a baseball team. Since a baseball team has nine members, it affords an excellent opportunity to establish a bi-univocal multiplication of two relationships.

To facilitate identification of each paragraph, a code will be used. Each item of the test will have a four-letter descriptor, the first two letters designating the type of paragraph that precedes it, and the last two letters the type of question. The four paragraph descriptors are:

(1) CA—additive classification
(2) CM—multiplicative classification
(3) SA—additive seriation
(4) SM—multiplicative seriation

The descriptors illustrate the dichotomous nature of the paragraph categories. In effect, they form a 2 x 2 matrix, set up to test the validity of the fourth and fifth questions (see Questions in Chapter III).

The Questions. Like the paragraphs, the questions asked are divided into two dichotomous categories, and attempt to test questions six and seven. The categories and descriptors for the test questions are:

(1) EI—easy, intension-based
(2) EX—easy, extension-based
(3) HI—hard, intension-based
(4) HX—hard, extension-based
An "easy" item is one judged easy by criterion (a) (number of logical steps between assumptions in additive hierarchies) or criterion (b) (number of properties or ordered difference scales in multiplicative hierarchies). A "hard" item is one judged difficult by the same two criteria. Figure 8 shows easy and hard items for the CA (additive classification) paragraph.

### ITEMS

**EASY**
- Betty, Dan and Brad
  - a) have pets.
  - b) are in the fifth grade.
  - c) like ice cream.*
  - d) do not have pets.
  - e) No answer is right.

**HARD**
- Dan and Brad
  - a) are in the fourth grade.*
  - b) do not have pets.
  - c) have dogs.
  - d) do not like ice cream.
  - e) No answer is right.

Where:
- A = children who go to Elm St. School
- B = fourth-graders (Betty, Dan, Brad, Karen)
- C = ice cream lovers (Betty, Dan, Brad)
- D = pet owners (Dan, Brad)

**DIGRAPHS**

Fig. 8. An easy and a hard item as judged by criterion (a) from Paragraph CA of the experimental test. The asterisk* denotes correct responses.

Items that emphasize intension present the property (of a class) or the relation (of a series) in the stem of the question. Items that emphasize extension present or list the elements (objects or classes) of the class or series in the stem of the question. Figure 9 illustrates
this dichotomy with two items from Paragraph CA.

EMPHASIS ON INTENSION:

Who likes ice cream?
  a) Dan, Betty and Brad
  b) Greg, Karen and Betty
  c) Brad, Betty and Karen
  d) Betty, Greg and Brad
  e) No answer is right.

EMPHASIS ON EXTENSION:

Betty, Dan and Brad
  a) have pets.
  b) are in the fifth grade.
  c) like ice cream.
  d) do not have pets.
  e) No answer is right.

Fig. 9. Sample intension and extension-based items from Paragraph CA.

Every item that is on the experimental test is thus identified by a four-letter descriptor. As an example, the two items in Figure 9 would be labelled CAEI and CAEX, since they both follow an additive classificatory paragraph, are judged easy, and emphasize intension in the first case (labelled "I") and extension in the second case (labelled "X"). Figure 10 shows the matrix of items for the entire test. There are sixteen possible combinations of item types, with two of each type of item, for a total of 32 items for the entire test. This arrangement becomes more important as the experimental design of the study is considered.

Forms. Because the order of presentation of either the paragraphs or the questions could affect the subjects'
responses, three forms—A, B and C (see Appendixes)—were prepared. All forms contained the same paragraphs, and all paragraphs were followed by the same eight questions. But the order of presentation of both paragraphs and questions were determined by reference to a table of random numbers (Dayton, 1970, pp. 379-383).

Administration Procedures

The test was administered to both third-graders and sixth-graders in one sitting, at approximately nine o'clock in the morning. The sixth-graders were tested on March 1, 1972, and the third-graders on March 9, 1972. Two intact third grades took the test in their classrooms. The sixty sixth-graders, because they were former students of Prairie Lincoln Elementary School, were directed to report to the cafeteria where the test was administered to them as a whole group. Both groups followed the procedure outlined

Fig. 10. Matrix of item descriptors for the experimental test.
(a) Tests were stacked so that forms alternated, i.e., Form A, Form B, Form C, Form A, etc., and were distributed to the subjects in this fashion.

(b) Subjects were told that the test results would not affect their grades, and that the investigator was trying to find out "how boys and girls in the third (or sixth) grade answer questions on a test like this."

(c) Subjects were given as much time as they needed to finish the test, and were told that the proctors could pronounce for them any word that they thought they did not know.

Statistical Analysis

Data sources. There were four sources of data used in this study:

(a) scores from the experimental test;

(b) IQ scores from the Lorge-Thorndike Intelligence Tests (for both third and sixth grades);

(c) reading vocabulary and reading comprehension test scores from the Iowa Silent Reading Test (for the third-graders only);

(d) reading comprehension test scores from the Paragraph Understanding subtest of the Monroe-Sherman Group Diagnostic Reading Aptitude and Achievement Tests.
**Relationships.** Two measures of relationships were needed, a measure of test reliability and a measure of relationships between group mean scores on categories of items on the experimental test and standardized test data.

Usual split-half methods of estimating reliability were discarded for reasons outlined earlier (see RELATED RESEARCH). Instead, an estimate of reliability based upon item statistics was employed. Since it was expected that certain items would be intrinsically more difficult than others, Horst's modification of the Kuder-Richardson Formula 20 (Guilford, 1965, p. 461) was selected.

Pearson product-moment correlation coefficients were obtained for an estimate of relationships between mean scores of third and sixth-graders on question categories of the experimental test and mean scores of both groups on the standardized IQ and reading tests. To obtain these statistics, the IBM 360 computer at the Ohio State University Instruction and Research Computer Center was used, programmed with version 52.5 of the P-STAT program for job statistics.

**Differences.** The primary purpose of this research was to assess differences, if any, among four dichotomous categories of items on the experimental test, i.e., classificatory versus serial, additive versus multiplicative, easy versus hard, and intension versus extension. It was also necessary to compare performance of third-graders
and sixth-graders on all categories and total test performance. Since each subject answered every item, subjects served as their own controls and a repeated-measures analysis of variance was the logical choice to test for significant differences. The IBM 360 computer at the University Computer Center was also used for this analysis.

Specifically, the study used a 2 x 8 repeated-measures design, a two-factor mixed model with one between-groups and one within-groups variable. The third and sixth-grades were the two levels of the between-groups variable, and the eight (four dichotomies) question categories were the repeated measures, i.e., the eight levels of the within-groups variable. The investigator was aware that the repeated measures were not entirely independent of one another, since they consisted of items that were classified four ways, and thus represented four "shufflings" of the total thirty-two test items into eight categories so that comparisons can be made.

This is defensible, however, since the design of the experiment is such that for every dichotomy (e.g., serial versus classificatory items) the effects of the other three dichotomies have been equally distributed among the two groups it items in each dichotomy. For example, examine the item descriptors for the classificatory-serial dichotomy shown in Figure 11.
### CLASSIFICATORY ITEMS
- CAEI
- CAEX
- CAHI
- CAHX
- CMEI
- CMEX
- CMHI
- CMHX

### SERIAL ITEMS
- SAEI
- SAEEX
- SAHI
- SAHX
- SMEI
- SMEX
- SMHI
- SMHX

**Fig. 11.** Item descriptors for items in classificatory and serial categories. In the actual test, there are two items of each type denoted by a descriptor.

While the 16 items (two for each descriptor) in each category differ with respect to the first descriptor (C and S), the remaining three descriptors are identical for each item. Thus the effects of each of the other categories are equally distributed within the classificatory and serial categories, and the comparison is legitimate.

However, only 16 comparisons of this type are possible. These are: (a) the four question dichotomies within the third-grade scores, (b) the four question dichotomies within the sixth-grade scores, and (c) the third-sixth grade level comparisons for each of the eight groups of items, and the total test score. There is no justification for comparing, for example, serial items with extension items, since the extension items category includes both serial and classificatory items, while the former group obviously does not. So the analysis of variance for repeated measures is justifiable as long as the post hoc comparisons are confined
to the 16 previously listed.

It may appear that the multiple use of test items in a repeated measures design yields too few comparisons for the effort invested. For this research, only 16 pairwise comparisons out of a possible 128 are allowed. But there is also considerable to be gained from such an arrangement. Dayton (1970) cites the two major drawbacks of repeated measures designs: (a) the passage of time between repeated measurements may significantly alter the subjects' performance, and (b) the possibility of multiple-treatment interference. In this study, the first drawback is eliminated, since all data from each measurement was gathered in a single administration. The second drawback is not prevented but equalized by constructing three forms of the test with separate random orders of paragraphs and questions.

Although additional comparisons beyond the 16 listed are not logically permissible, they still exist as potential comparisons within the structure of the analysis of variance. Since 16 is but a small proportion of 128 potential contrasts Dunn's procedure for post hoc comparisons can be used. The alpha risk, set at .05 for the entire experiment, is thus divided 16 ways and each post hoc comparison will be made at the .003 level.

The final difference comparison to be made is between mean total scores on the experimental test of third and sixth-graders. Since the total possible score on the entire
test is 32, and the highest possible score on any of the
eight item categories is 16, a separate t-test will assess
the significance of the difference between these two means.

In summary, there are a total of 12 (11 for the sixth-
grade group) sources of data in this research. They are:

Standardized test means:
(1) IQ
(2) reading vocabulary (for third-graders only)
(3) reading comprehension

Experimental test means:
(4) items following classificatory paragraphs
(5) items following serial paragraphs
(6) items following additive paragraphs
(7) items following multiplicative paragraphs
(8) easy items
(9) hard items
(10) intensibility-based items
(11) extension-based items
(12) total test score

Correlations between all variables were obtained, and
reliability of the experimental test was estimated by use of
Horst's modification of Kuder-Richardson Formula 20.

Differences between means of groups within four
dichotomous item categories of the experimental test were
tested by use of a repeated measures analysis of variance.
CHAPTER V

RESULTS

After presenting the general descriptive statistics, each research question stated in Chapter III will be restated, and evidence presented for its explanation and interpretation.

General Descriptive Statistics

As was mentioned in Chapter IV, the Horst modification of the Kuder-Richardson Formula 20 for internal consistency reliability was computed for both the third and sixth-grade groups of scores. Reliability for the third-grade testing was .736, and for the sixth-grade testing was .837. Although these figures were not acceptable by normal testing standards, they were sufficient for an experimental instrument administered for the first time.

Table 2 shows the means and standard deviations of both groups for IQ, reading vocabulary (third grade only), reading comprehension and all question categories and total performance on the experimental test. The IQ scores are extremely homogeneous, the means for both grade levels
within three points of each other, and standard deviations nearly identical. Standardized reading test scores show both grades testing at about grade level. In general, both groups tended to be relatively homogeneous, which would add support to any differences due to type of question on the experimental test.

**TABLE 2**

Means and Standard Deviations for all Subjects on Standardized and Experimental Tests

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>Test Scores</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade Test Scores</td>
<td>Voc.</td>
<td>Read.</td>
<td>Com.</td>
</tr>
<tr>
<td>Third Mean S.D.</td>
<td>102.21</td>
<td>3.42</td>
<td>3.71</td>
<td>7.73</td>
</tr>
<tr>
<td>Sixth Mean S.D.</td>
<td>105.23</td>
<td>0.95</td>
<td>1.11</td>
<td>3.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Third Mean S.S.</td>
<td>6.61</td>
<td>8.53</td>
<td>5.81</td>
<td>7.45</td>
</tr>
<tr>
<td>Sixth Mean S.D.</td>
<td>9.20</td>
<td>11.90</td>
<td>9.07</td>
<td>10.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Hard Ques.</th>
<th>Intens. Ques.</th>
<th>Extens. Ques.</th>
<th>Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third Mean S.D.</td>
<td>6.90</td>
<td>6.96</td>
<td>7.38</td>
<td>14.35</td>
</tr>
<tr>
<td>Sixth Mean S.D.</td>
<td>10.43</td>
<td>10.20</td>
<td>10.62</td>
<td>20.80</td>
</tr>
</tbody>
</table>

**Notes:**

a Only scores for third-graders were available.
b Total possible score was 16 for all categories.
c Total possible score was 32.

In both the third and sixth-grade groups, performance on the experimental test was consistent. The total possible test score was 32, and the scores ranged from 4 to 25 for the
third-graders, 7 to 29 for the sixth-graders. Performances on the individual question categories was also fairly even within each grade level, and these differences will be discussed fully within the context of each individual research question.

Questions: Relationships

Will there be significant relationships between experimental test performance and IQ as measured by traditional instruments?

Tables 3 and 4 show all correlations between standardized and experimental test scores for both groups of subjects. In the case of IQ scores, what is immediately noticeable is the differences between relationship of the experimental test to IQ, when compared to that of the standardized reading tests. Both third and sixth-grade scores on reading comprehension correlate highly (.73 and .74 respectively) with IQ as measured by the Lorge-Thorndike Intelligence Test, but this relationship drops to .43 for the third-graders and .59 for the sixth-graders for total scores on the experimental test.

Examining relationships between IQ and scores on individual question categories, the correlations drop even further, particularly in the case of the third-graders. The correlation between performance on serial paragraphs and IQ does not even reach significance in the case of the third-graders, although it is one of the highest of the
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ</td>
<td>.68**</td>
<td>.73**</td>
<td>.47**</td>
<td>.20</td>
<td>.44**</td>
<td>.28*</td>
<td>.29*</td>
<td>.53**</td>
<td>.29*</td>
<td>.45**</td>
<td>.43**</td>
</tr>
<tr>
<td>R. Voc.</td>
<td>.83**</td>
<td>.50**</td>
<td>.49**</td>
<td>.32*</td>
<td>.42**</td>
<td>.54**</td>
<td>.37**</td>
<td>.53**</td>
<td>.51**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. Comp.</td>
<td>.51**</td>
<td>.35**</td>
<td>.51**</td>
<td>.40**</td>
<td>.42**</td>
<td>.60**</td>
<td>.41**</td>
<td>.54**</td>
<td>.54**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classif.</td>
<td>.36**</td>
<td>.74**</td>
<td>.76**</td>
<td>.79**</td>
<td>.82**</td>
<td>.75**</td>
<td>.79**</td>
<td>.87**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serial</td>
<td>.74**</td>
<td>.57**</td>
<td>.75**</td>
<td>.66**</td>
<td>.68**</td>
<td>.69**</td>
<td>.77**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addit.</td>
<td></td>
<td>.48**</td>
<td>.83**</td>
<td>.82**</td>
<td>.75**</td>
<td>.84**</td>
<td>.90**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi.</td>
<td></td>
<td></td>
<td>.78**</td>
<td>.73**</td>
<td>.75**</td>
<td>.71**</td>
<td>.82**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td></td>
<td></td>
<td></td>
<td>.70**</td>
<td>.83**</td>
<td>.84**</td>
<td>.94**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.77**</td>
<td>.83**</td>
<td>.91**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inten.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.58**</td>
<td></td>
<td></td>
<td>.87**</td>
</tr>
<tr>
<td>Exten.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.91**</td>
</tr>
</tbody>
</table>

*P < .05  
**P < .01
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ</td>
<td></td>
<td>.74**</td>
<td>.48**</td>
<td>.57**</td>
<td>.54**</td>
<td>.52**</td>
<td>.54**</td>
<td>.56**</td>
<td>.57**</td>
<td>.52**</td>
<td>.59**</td>
</tr>
<tr>
<td>R. Comp.</td>
<td></td>
<td>.38**</td>
<td>.53**</td>
<td>.49**</td>
<td>.50**</td>
<td>.42**</td>
<td>.54**</td>
<td>.46**</td>
<td>.50**</td>
<td>.52**</td>
<td>.52**</td>
</tr>
<tr>
<td>Classif.</td>
<td></td>
<td>.55**</td>
<td>.75**</td>
<td>.73**</td>
<td>.88**</td>
<td>.78**</td>
<td>.79**</td>
<td>.84**</td>
<td>.89**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serial</td>
<td></td>
<td>.79**</td>
<td>.70**</td>
<td>.74**</td>
<td>.88**</td>
<td>.83**</td>
<td>.79**</td>
<td>.84**</td>
<td>.87**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addit.</td>
<td></td>
<td>.50**</td>
<td>.82**</td>
<td>.81**</td>
<td>.77**</td>
<td>.85**</td>
<td>.87**</td>
<td>.87**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi.</td>
<td></td>
<td>.71**</td>
<td>.80**</td>
<td>.77**</td>
<td>.73**</td>
<td>.81**</td>
<td>.92**</td>
<td>.92**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy</td>
<td></td>
<td></td>
<td>.74**</td>
<td>.88**</td>
<td>.83**</td>
<td>.92**</td>
<td>.92**</td>
<td>.92**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard</td>
<td></td>
<td></td>
<td></td>
<td>.84**</td>
<td>.90**</td>
<td>.94**</td>
<td>.94**</td>
<td>.94**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inten.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.71**</td>
<td>.92**</td>
<td>.92**</td>
<td>.92**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exten.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.93**</td>
</tr>
</tbody>
</table>

*p < .05
**p < .01
sixth-grade correlations. This low level of relationship is not confined to any particular category of question, but is uniform throughout all categories and total test score. The range is a bit wider (.20 to .53) in the third-grade group than in the sixth-grade group (.48 to .59), but this is to be expected with tests given to younger children.

What seems significant is that even the highest single correlation between the Lorge-Thorndike group IQ test and the experimental test, the total test score of the sixth-grade group (.59), shares only about 36 per cent of the variance with the group IQ test. This same correlation in the case of the third-graders (.43) accounts for only 18 per cent. Since group IQ tests require a certain amount of reading, undoubtedly much of this shared variance is due to word recognition and vocabulary ability. The difference between the third-graders' and sixth-graders' estimates of shared IQ test and experimental test variance would seem to support this contention.

The fact that within each grade level (a) correlations between the experimental test and IQ are lower than those between standardized reading measures and IQ, and (b) both the total test score and individual categories of questions' correlations are consistent within each grade level (especially in the case of the sixth-graders), supports a tentative conclusion that the experimental test is measuring an aspect of reading that is not measured on standardized
reading tests, and is not reflected in performance on standardized group IQ tests. This could possibly be the ability that the experimental test was constructed to measure: thinking applied to reading.

Will there be significant relationships between experimental test performance and reading vocabulary or reading comprehension scores on traditional instruments?

Reading vocabulary scores were available for the third-graders only. They were obtained from the reading vocabulary subtest of the Iowa Silent Reading Test. Aside from the .83 correlation with the comprehension subtest of the same test, sections and total scores of the experimental test related little better than IQ scores did. Correlations range from .32 to .54, and are extremely close to the reading comprehension coefficients. The scores for the third-graders on the comprehension subtest of the Iowa Silent Reading Test range from .35 to .60. It is remarkable to compare the correlations between these two subtests with the experimental test variables, as the coefficients are nearly identical. For reading vocabulary and comprehension respectively, the correlations between classificatory paragraphs are .50 and .51, between these test scores and serial paragraphs, .49 and .51, multiplicatives, .39 and .40, easy questions, .42 in both cases, hard questions, .54 and .60, intension questions, .37 and .41, extension items, .53 and .54, and total test score, .51 and .54.
Because of the high correlation between the reading vocabulary and comprehension subtests, and the uniformly low correlations between all sections and total score of the experimental test, it seems reasonable to conclude that the reading comprehension subtest of the Iowa Test is largely measuring reading vocabulary rather than reasoning in reading in the case of the third-graders. These two subtests share 69 per cent of their total variance, while accounting for only 25 per cent of the variance on the experimental test. Both standardized subtests also relate highly to the group IQ test, which, as has been shown, is even less related to any of the experimental test variables.

A glance at the correlations between experimental test variables and reading comprehension subtest score on the Monroe-Sherman Reading Test for the sixth-graders is even more compelling. Even though the scores were obtained from a much older group of children on an entirely different instrument than that used for the third-graders, the correlations remain consistent with those in the third-grade group, varying only slightly from category to category. In fact, correlations between total experimental test performance and standardized reading vocabulary (for third-graders only) and reading comprehension (for both groups) are very close: .51, .54 and .52 respectively.

All evidence so far underscores what was stated in the review of related research for this study, that standardized
reading comprehension subtests lack validity, that they measure hardly more than reading vocabulary tests, and that they do not measure ability to reason in reading. There is little to support a contention that general intelligence contributed much to the gap between the experimental test performance and standardized reading measures, for IQ scores also correlated only slightly with the experimental test scores. While it is normal for IQ and reading test scores to correlate more highly in the upper grades, this trend is only partially true in the case of the experimental test. The correlations between IQ and experimental test performance indicate a shared variance of only about 25 per cent at the sixth-grade level.

It therefore seems logical to conclude that the experimental test is measuring an ability that is measured only slightly by standardized IQ, reading vocabulary and reading comprehension tests.* Since the test items were constructed from tasks that Inhelder and Piaget (1964) found to be especially relevant to children’s thinking from the ages of seven to eleven, and since there was no opportunity for subjects to use concrete props of any

*It is also possible to conclude that the standardized IQ and reading instruments in this study were simply measuring the same abilities. Nevertheless, it is still apparent that the experimental test is only slightly related to any of these measures.
kind, the experimental test most probably is a measure of
cognitive ability applied to a reading medium—reasoning
in reading.

Before proceeding to differences, an examination of
correlations between the various question categories of
the experimental test is warranted. The overwhelming
impression is one of consistency and stability, especially
within the sixth-grade group. Within the third-graders,
coefficients range from .35 to .84 (with the majority in
the .60 to .80 range) for relationships between various
question categories, and from .77 to .94 between question
categories and total test performance. For the sixth-
graders, question categories correlate .50 to .90 among
themselves, and .81 to .94 with total test performance.

The consistency of intercorrelations between test
sections reflects consistent mean score performance on
these sections, which leads to the question of differences.

Questions: Differences

**Will there be significant differences between experimental
test performance of third-graders and sixth-graders?**

To answer this and the next five questions, a repeated
measures analysis of variance was performed on the data
(see Chapter IV for details of design). It will be recalled
that Table 2 showed that mean scores on all question
categories of the experimental test were not extremely
deviant. This data has been condensed into the form shown in Table 5.

### TABLE 5

Mean Scores for Question Categories

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Third</td>
<td>7.7</td>
<td>6.6</td>
<td>8.5</td>
<td>5.8</td>
<td>7.5</td>
<td>6.9</td>
<td>7.0</td>
<td>7.3</td>
<td>14.4</td>
</tr>
<tr>
<td>Sixth</td>
<td>11.6</td>
<td>9.2</td>
<td>11.9</td>
<td>9.0</td>
<td>10.4</td>
<td>10.4</td>
<td>10.2</td>
<td>10.6</td>
<td>20.8</td>
</tr>
</tbody>
</table>

Note—All scores have been rounded to one decimal.

Thus it was not surprising to find that, as shown in Table 6, the F ratios were only significant for the grade level effect. Since the between groups (grade level) variable was the only significant effect, it was not necessary to use Dunn's procedure or any other post hoc test. All differences between third and sixth-grade scores were clearly significant. Total score was also
assumed significant, although a $t$-test was run out of curiosity. This difference was significant at the .01 level ($t = 6.89$, 59 df). Box's correction (Dayton, 1970) was performed, and the $F$ remained highly significant at the conservative degrees of freedom. Thus heterogeneity of covariance was small enough to be ignored.

Will there be significant differences between subjects' performance on any of the four dichotomous item categories, or any interaction between grade level and item categories?

Because of the lack of significant $F$ ratios in the treatment (item category) and interaction sources, it can be concluded that no significant differences were found for any of the predicted categories in Questions Three through Eight stated in Chapter III.

The absence of significant differences between subjects' performance on classification and seriation paragraphs or additive and multiplicative paragraphs supports Inhelder and Piaget's (1964) theory that these two cognitive abilities develop simultaneously in the period of concrete operations, even within a reading setting. The lack of significant differences between easy and hard question categories with-draws support from the establishment of difficulty in terms of either the number of logical steps as represented by lines in a digraph between objects or classes in additive paragraphs, (criterion a) or the number of properties or difference scales used to refer to a class or object in multiplicative paragraphs
(criterion b). The notion is also dispelled that multiple-choice items based on the intension or extension of a class or series established in a paragraph present tasks of differential ease.

Summary

The experimental test showed an internal consistency reliability of .736 and .837 for administrations given to third-graders and sixth-graders (N = 60 in each group). Inspections of the means and standard deviations of standardized group IQ, reading vocabulary (for the third grade only) and reading comprehension tests indicated that the groups were fairly homogeneous with respect to IQ and reading grade level, given expected age differences. Experimental test means reflected this homogeneity.

Correlations between standardized tests were high for both grades. Correlations between the standardized tests and all categories of the experimental test, however, were uniformly low for both grades. Correlations among the sections of the experimental instrument were high.

Sixth-graders outperformed third-graders on all parts of the experimental test, and these differences were significant. There were no significant differences, in either the third or the sixth-grade group, between item categories, and no significant interactions.
CHAPTER VI

SUMMARY AND CONCLUSIONS

Summary

Reading comprehension is the sine qua non of the entire reading process. Without comprehension, reading is meaningless word-calling, similar to the choir boy's chanting of Latin verses that are pleasant to his ear but devoid of interpretation unless a translation is made available to him. Essential as this ability is, no one seems to know exactly what occurs to bring about understanding of printed text in the mind of the reader.

After a review of the literature concerning the measurement of this important process revealed striking deficits with respect to validity and reliability in the standardized instruments presently used, it was decided to construct a new instrument based on one theory of child development that was at the same time heavily weighted toward cognitive rather than affective phenomena, and based upon a substantial body of empirical observation. Jean Piaget's developmental construct of intelligence fulfilled both these requirements.
The investigator compiled several methods of quantitatively examining prose paragraphs for evidences of Piagetian operations, particularly those of the concrete operational period, or the period of intellectual growth characterized by classification and seriation abilities in the presence of concrete props, which extends from ages seven to eleven in normal children.

Because Piaget had stated that concrete operations were beyond the scope of the concrete operational child when the child was deprived of concrete props, a pilot study was performed in which simple classification and seriation situations were presented to a group of 45 fourth-graders. This group, despite lower than average scores on standardized reading measures, showed clearly that they were able to apply these cognitive abilities to printed matter.

A 32-item multiple-choice test over four short paragraphs was then constructed. The four paragraphs represented classificatory-additive, classificatory-multiplicative, serial-additive and serial-multiplicative operations. Eight questions followed each paragraph, and represented four categories: easy versus hard, and intension-based versus extension-based. "Easy" and "hard" items were classified according to two criteria, (a) the logical distance between classes in a hierarchy or objects or classes on an ordered difference scale in additive
paragraphs, and (b) the number of class properties or ordered difference scales involved in identifying objects or classes in multiplicative paragraphs. Paragraphs and questions were presented in random orders on three forms of the test, and on each form each paragraph had four types of each question. This was possible due to the multiple classification of each item.

The test was administered to 60 third-graders in a predominately upper-middle class area elementary school, and to 60 sixth-graders who had previously attended the same elementary school. Scores on the Lorge-Thorndike Intelligence Test were obtained for all subjects. Scores on the Iowa Silent Reading Test in reading vocabulary and reading comprehension were obtained for the third-graders, and Monroe-Sherman comprehension subtest scores were obtained for the sixth-graders. Means and standard deviations of both groups on the standardized instruments revealed that they were extremely homogeneous given chronological age differences.

Experimental test data was reported in two ways: by relationships and by differences. To assess relationships, Pearson product-moment correlations were compiled for all subjects comparing standardized test variables with experimental test variables. Standardized test variables of IQ, reading vocabulary and reading comprehension correlated highly with each other, but to a very limited degree with
all of the experimental test variables. These correlations were slightly higher in the sixth-grade group, but still remained substantially below standardized test intercorrelations. Experimental test variables correlated very highly with one another, and internal consistency reliability coefficients of .736 and .837 were obtained, for the third-grade and sixth-grade administrations respectively.

Differences were assessed by use of a 2 x 8 repeated measures analysis of variance for sections of the experimental test. Sixth-graders scored significantly higher than third-graders on all item categories, but there were no significant differences between item categories in either group, and no significant interaction.

Conclusions

Reasoning in reading. The now-common phrase, "reasoning in reading," was coined by Edward Lee Thorndike in his classic 1917 study of mistakes in paragraph reading. Fifty-five years later, such little use has been made of his insight that Wayne Otto, Principal Investigator at the Wisconsin Research and Development Center for Cognitive Learning, was led to state in a recent article (Otto, 1971) that "Unfortunately, many of the very questions Thorndike left unanswered in his report remain unanswered—or unanswerable in any definitive way—to this day p. 440 ." Even after Husbands and Shores (1950) plea for measures of
problem-solving in reading, little was done to follow up what seemed to be an obvious need in the search for answers to the complex processes that are involved in understanding printed text.

It was this insight of Thorndike's, that reading comprehension involved the application of thinking skills to printed materials, that led the investigator to perform this research. All that was needed was a practical theory of cognitive development that was based upon observations of real children as they attempted to solve problems that required a variety of thinking skills. The theory chosen was that of the Swiss psychologist, Jean Piaget. Piaget admitted the futility of reaching for solutions to behavioral mysteries by recourse to neurological or physiological theories at the present time. He states:

Needless to say, it still remains a reasonable question to ask what are the psycho-physiological mechanisms underlying the co-ordination of actions, the retroactive and anticipatory adjustments and the operations themselves. But this would be a matter of explaining how these things become possible. What we have attempted to do is something very different, and that is to give a detailed account of the reason why, given the fact that they are physiologically possible, they assume these particular forms as and when they do [Inhelder and Piaget, 1964, p. 293].

This is precisely the goal of the study reported in this paper. It is hoped that this research would be the beginnings of a collection of empirical observations of thinking applied to reading, considered within the context
of child development. Even if such a mass of data leads nowhere in explaining the basic neurological processes involved in reading comprehension, it will help to clarify exactly what children of different cognitive abilities and ages are capable of doing when confronted with certain logical problems in printed form. Such information is badly needed by all reading researchers and test constructors, as well as by teachers who need to make definitive decisions as to what activities their children will engage in to further their understanding of what they read.

The data from this research led the investigator to two general conclusions:

(a) children can and do apply cognitive operations of classification and seriation that Piaget refers to as "concrete operations" in the reading setting without the addition of concrete props. Third-graders perform at about a 45 per cent success level, and sixth-graders perform at about a 65 per cent level. In addition, they deal with additive and multiplicative hierarchies equally well, the situation which prevailed in Piaget's experiments. Changing the forms of questions which follow the paragraphs did not seem to affect performance, which also agrees with Piaget's findings in the case of manipulation of concrete props.

(b) Standardized group intelligence, reading vocabulary and reading comprehension tests of the type used in this
study bear little relationship to the type of reading skills required by the experimental test, i.e., the application of classification and seriation tasks to simply-worded, short paragraphs constructed to present these hierarchies in situations familiar to most elementary school children.

Some supplementary data. The multiple classification paragraph of the experimental test was written by Thorndike and used in his 1917 study. Its inclusion seemed to make little difference, since performance on the items following this paragraph was comparable to performance on items that came after the other three paragraphs. Thorndike's question, "Who had red hair?" in multiple-choice form, ranked 12 in difficulty (out of 32) in both the third and sixth-grade testings. Third-graders answered it correctly 37 per cent of the time, and sixth-graders answered it correctly 63 per cent of the time, which placed performance on this item roughly equal to performance on the total test, which was an average of 45 per cent for the third grade, and 65 per cent for the sixth grade, as has been mentioned before.

To gain some idea of how the subjects felt about the experimental test, a three-choice opinion question was presented immediately following the test, i.e., on the last page of the instrument (See Appendixes). It asked if they thought the test was "easier," "harder" or "about the same"
as other tests they have taken. Most of the subjects seemed to think that the experimental test was about the same difficulty as other tests they normally take. Third-graders marked "about the same" 66 per cent of the time, "harder" 24 per cent of the time, and "easier" only 10 per cent of the time. The sixth-graders percentages for the same questions were 60, 4, and 36 respectively, indicating that they thought the test easier more often, which would be a logical occurrence as their scores were higher.

**Limitations of the study.** Before generalizations for future research can be made, there are several limitations to this study that should be presented.

(a) The paragraphs used on the test, while presenting classification and seriation hierarchies in a straightforward manner, were specially constructed for use in this study (except for Thorndike's paragraph, and it also was constructed by an investigator for a particular purpose). Thus they may not represent the typical day-to-day reading material encountered by third and sixth-grade children. Textbook reading would more closely resemble the reading done on the experimental test, although textbook paragraphs are seldom so easily worded.

(b) The subjects in the study were an extremely homogeneous group with respect to socioeconomic background, dialectical differences and cultural influences. Thus the results of this study cannot justifiably be generalized
outside of this group. In addition, results cannot be
generalized outside of the particular grade levels involved,
since both experimental test variables and grade levels
were fixed, rather than chosen at random.

(c) The construction of the experimental test,
although it involved considerable effort, was essentially
a first attempt to replicate Piaget's experiments in the
reading setting with a multiple-choice test. As such,
there is most probably a loss in accuracy in the trans-
lation of concrete manipulations to printed text, since, of
course, this translation is literally impossible. Also,
various typographical errors in making the final copies of
the test made it possible for two items to have two correct
answers. This occurred on items I-4 and II-5 of Form A
(items I-4 and II-1 of Form C), and item II-7 of Form B
(items I-4 of Form and II-1 of Form C). The actual wordings
of these items can be found in Appendixes C, D and E of this
paper.

(d) The data analysis, while testing for both relation-
ships and differences, could be supplemented to reveal much
more information than was presented in this report. These
additional analyses will be recommended for future research
efforts.

Implications for future research. While the correla-
tions presented in Chapter V seem to support the contention
that the experimental test is measuring a unique aspect of
reading comprehension, a factor analysis of the data could pinpoint the contributions of abilities measured on the standardized tests. This would certainly be the first action undertaken to supplement the data revealed in this study.

Because differential construction of sentences in the experimental test paragraphs or questions may have played a significant role in the subjects' comprehension of them, a syntactic analysis of the data is probably warranted. Items could be arranged according to their dominant syntactical construction, and reanalyzed for significant differences, as could the paragraphs. Cloze tests over the paragraphs, with certain grammatical items deleted, might also be a possibility in this area.

Finally, the paragraphs might be individually presented to subjects in an interview situation, similar to Piekarz's (1956) research. This would also be in line with the "clinical method" of research used by Piaget and his associates for compiling the data used as a basis for this study. This final method might be the most productive of all future research efforts.

For years the term "developmental reading" has been applied to the normal progress that children make in reading as they proceed through the school years and into adulthood. Unfortunately, most definitive models of developmental reading are not developmental at all, but based
upon research evidence gathered on groups of older subjects who are already competent readers, or upon the abnormal behavior of those individuals who are seriously retarded in reading.

Now that it has been shown that children of elementary school age can apply classification and seriation abilities to data entirely in printed format, further research efforts are needed to compile a data bank which will form the foundation for the first truly developmental theory of reading behavior, one which bases its generalizations upon empirical observations of children at different stages of mental development as they apply their growing mental skills to reading. Only from such a theory can come the greatest amount of immediately useful techniques for measuring the ability to reason in reading, and fostering its continued growth in the classroom.
REFERENCES


Davis, F. B. Fundamental factors of comprehension in reading. Psychometrika, 1944, 9, 185-197.


Davis, J. E. The ability of intermediate grade pupils to distinguish between fact and opinion. The Reading Teacher, 1969, 22, 419-422.


Farr, R. Reading: what can be measured? Newark, Delaware: International Reading Association, 1969.


Fraser, L. T. Structural analysis of the knowledge that results from thinking about text. Journal of Educational Psychology Monograph, 1969, 60, 1-16.

Gans, R. A study of critical reading comprehension in the intermediate grades. Teachers College Contributions to Education, 1940, No. 811.


APPENDIX A: Pilot Study Test

Name___________________________Age____Boy or Girl?_______

DIRECTIONS: Read each story carefully. Then write the answer to each question in the space under the question. Take as much time as you want.

Jim runs faster than Ted.
Dan runs slower than Ted.

1. Who runs the fastest?
2. Who runs the slowest?

Mary was buying Christmas presents. She saw a ring that cost more than a pen. Then she looked at a bracelet that cost less than the pen. After a while, she found a necklace that cost more than the ring.

3. Which present cost the most?
4. Which present cost the least?
5. Which present cost more than the bracelet, but less than the ring?
6. Which present cost more than the ring and the pen?

My uncle lives farther away than my grandfather.
My cousin lives closer than my aunt, who lives closer than my grandfather.

7. Who lives closest to me?
8. Who lives farthest away from me?
9. Who lives closer than my grandfather, but farther away than my cousin?

10. Who lives farther away than my aunt, but closer than my uncle?

In a far-away country lived some strange people. In this country there were green people and yellow people. Green people lived in the hills and the valleys. Yellow people lived in the mountains or by the river. The mountain people were shepherds and tool-makers. Valley people were builders or drivers. The hill people were either farmers or hunters. Those who lived by the river were doctors or teachers.

11. Where would a teacher live?

12. What kinds of jobs could hill people have?

13. Are there more green people or strange people?

14. What color could a tool-maker be?

15. Where could yellow people live?

16. Are there more builders or more valley people?

17. What color could a hunter be?

18. What color are mountain people who are not shepherds?

19. What jobs do the people have who are not green and do not live by the river?

20. Are there more people who are not drivers or more people who do not live in the hills?
### Mean Test Scores

<table>
<thead>
<tr>
<th>Subjects</th>
<th>N</th>
<th>Classification (1-10)</th>
<th>Seriation (11-20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>45</td>
<td>5.22 (s.d. = 1.69)</td>
<td>7.68 (s.d. = 2.25)</td>
</tr>
<tr>
<td>Boys</td>
<td>20</td>
<td>4.45</td>
<td>7.90</td>
</tr>
<tr>
<td>Girls</td>
<td>25</td>
<td>4.84</td>
<td>7.52</td>
</tr>
<tr>
<td>9-yr.-olds</td>
<td>15</td>
<td>5.91</td>
<td>7.73</td>
</tr>
<tr>
<td>10-yr.-olds</td>
<td>27</td>
<td>4.74</td>
<td>6.92</td>
</tr>
<tr>
<td>11-yr.-olds</td>
<td>3</td>
<td>7.00</td>
<td>8.30</td>
</tr>
</tbody>
</table>

### Item Difficulty

#### Classification Items

<table>
<thead>
<tr>
<th>Item</th>
<th>% correct responses</th>
<th>difficulty rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>98</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>93</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>87</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>56</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>44</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>80</td>
<td>6 or 7</td>
</tr>
<tr>
<td>7</td>
<td>62</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>80</td>
<td>6 or 7</td>
</tr>
<tr>
<td>9</td>
<td>69</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>67</td>
<td>4</td>
</tr>
</tbody>
</table>

#### Seriation Items

<table>
<thead>
<tr>
<th>Item</th>
<th>% correct responses</th>
<th>difficulty rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>67</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>73</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td>44</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>62</td>
<td>8</td>
</tr>
<tr>
<td>15</td>
<td>60</td>
<td>6 or 7</td>
</tr>
<tr>
<td>16</td>
<td>56</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>60</td>
<td>6 or 7</td>
</tr>
<tr>
<td>19</td>
<td>38</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>29</td>
<td>1</td>
</tr>
</tbody>
</table>
APPENDIX C: The Experimental Test--Form A

DIRECTIONS: Read each story, then circle the best answer.

I

I have five good friends. Bert lives right next door to me. Carol lives farther away than Bert. Jill lives even farther away than Carol. Dana lives farther away than Carol, but closer than Scott. I have to take a bus to his house.

1. Carol lives
   a) farther away than Jill and Bert.
   b) farther away than Dana and Jill.
   c) closer than Scott and Bert.
   d) closer than Bert and Jill.
   e) No answer is right.

2. Who lives farther away than Dana?
   a) Jill
   b) Bert
   c) Scott
   d) Carol
   e) No answer is right.

3. Who lives closest to me?
   a) Scott
   b) Jill
   c) Dana
   d) Bert
   e) No answer is right.

4. Carol
   a) lives closer than Dana.
   b) lives farther away than Jill.
   c) lives closer than Jill.
   d) lives farther away than Dana.
   e) No answer is right.

5. Dana lives farther away than
   a) Jill, Carol and Bert.
   b) Bert and Scott.
   c) Jill and Scott.
   d) Carol, Jill and Scott.
   e) No answer is right.
6. Who lives farthest away from me?
   a) Jill
   b) Carol
   c) Bert
   d) Scott
   e) No answer is right.

7. Who lives closer than Scott and Jill?
   a) Dana and Carol
   b) Carol and Bert
   c) only Carol
   d) only Bert
   e) No answer is right.

II

John had two brothers who were both tall. Their names were Will and Fred. John's sister, who was short, was named Mary. John liked Fred better than either of the others. All these children except Will had red hair. He had brown hair.

1. Fred and Will
   a) have brown hair.
   b) are not John's brothers.
   c) are tall.
   d) have red hair.
   e) No answer is right.

2. Who had red hair?
   a) Will, Fred and John
   b) John, Mary and Will
   c) only Fred and John
   d) Mary, John and Fred
   e) No answer is right.

3. Who is John's brother, is short, and has brown hair?
   a) Will
   b) Fred and Will
   c) John and Mary
   d) Will, Fred and John
   e) No answer is right.

4. Who is tall, is not John, and has red Hair?
   a) Fred and Mary
   b) Fred
   c) Will and Fred
   d) Will
   e) No answer is right.
5. Mary
   a) is short, is not John's brother, and has red hair.
   b) is tall, is not John's brother, and has brown hair.
   c) is short, is John's sister, and has red hair.
   d) is tall, is John's sister, and has red hair.
   e) No answer is right.

6. John and Mary
   a) have red hair.
   b) have brown hair.
   c) are tall.
   d) are short.
   e) No answer is right.

7. Which children are not John's brothers?
   a) Will and Mary
   b) Fred and John
   c) John and Mary
   d) Will and Fred
   e) No answer is right.

8. Will
   a) is short, is John's brother, and has red hair.
   b) is tall, is not John's brother, and has brown hair.
   c) is short, is not John's brother, and has red hair.
   d) is tall, is John's brother, and has brown hair.
   e) No answer is right.

III

Betty, Brad, Dan, Greg and Karen go to Elm St. School. Greg is in the fifth grade, and the others are in fourth. All the fourth-graders like ice cream except Karen. Betty does not have a pet. The other fourth-graders who like ice cream to have pets. Dan has a dog.

1. Who likes ice cream?
   a) Dan, Betty and Brad
   b) Greg, Karen and Betty
   c) Brad, Betty and Karen
   d) Betty, Greg and Brad
   e) No answer is right.

2. Who is not in the fourth grade?
   a) Karen
   b) Betty
   c) Dan
   d) Greg
   e) No answer is right.
3. Who goes to Elm St. School?
   a) only Karen, Brad, Betty and Greg
   b) only Dan, Betty, Greg and Brad
   c) Dan, Greg, Karen, Brad and Betty
   d) everyone but Greg
   e) No answer is right.

4. Betty, Dan and Brad
   a) have pets.
   b) are in the fifth grade.
   c) like ice cream.
   d) do not have pets.
   e) No answer is right.

5. Brad
   a) does not like ice cream.
   b) is in the fourth grade.
   c) has a pet cat.
   d) does not go to Elm St. School.
   e) No answer is right.

6. Betty and Brad
   a) don't like ice cream.
   b) have pets.
   c) are not in the fourth grade.
   d) do not have pets.
   e) No answer is right.

7. Who is not in the fifth grade?
   a) only Greg
   b) only Karen
   c) only Betty, Dan and Karen
   d) only Dan, Betty, Karen and Brad
   e) No answer is right.

8. Dan and Brad
   a) are in the fourth grade.
   b) do not have pets.
   c) have dogs.
   d) do not like ice cream.
   e) No answer is right.

IV

There are nine boys on the Bears baseball team. Mark, Bill and Ed, are all taller than Henry or Dave or Bob. Jack, Ray and Al are the shortest. Henry, Jack and Mark run faster than Dave, Ray or Bill. Bob, Ed and Al are the slowest runners.
1. Who is slower and taller than Dave?
   a) Jack
   b) Henry
   c) Bill
   d) Ed
   e) No answer is right.

2. Bill is taller and faster than
   a) Ed and Mark.
   b) Henry and Dave.
   c) Bob and Al.
   d) Ray and Jack.
   e) No answer is right.

3. Ray runs faster than
   a) Mark and Henry.
   b) Jack, Dave and Mark.
   c) Henry and Jack.
   d) Bill.
   e) No answer is right.

4. Who is taller than Dave?
   a) Ed and Mark
   b) Bill and Al
   c) Mark and Al
   d) Ray and Bob
   e) No answer is right.

5. Jack is faster and shorter than
   a) Bill and Ed.
   b) Henry and Mark.
   c) Dave and Ray.
   d) Al and Bob.
   e) No answer is right.

6. Who runs faster than Ray?
   a) Al and Jack
   b) Henry and Mark
   c) Mark and Dave
   d) Jack and Ed
   e) No answer is right.

7. Who is taller than Henry and faster than Ed?
   a) Dave and Bill
   b) Bill and Mark
   c) Bob and Bill
   d) Dave and Bob
   e) No answer is right.
8. Henry is taller than
   a) Ed.
   b) Al and Dave.
   c) Jack.
   d) Jack and Bob.
   e) No answer is right.

This test was
   a) harder than most tests I've taken.
   b) easier than most tests I've taken.
   c) about the same as most tests I've taken.

APPENDIX D: The Experimental Test--Form B

DIRECTIONS: Read each story, then circle the best answer.

I

John had two brothers who were both tall. Their names were Will and Fred. John's sister, who was short, was named Mary. John liked Fred better than either of the others. All these children except Will had red hair. He had brown hair.

1. Which children are not John's brothers?
   a) Will and Mary
   b) Fred and John
   c) John and Mary
   d) Will and Fred
   e) No answer is right.

2. Who is John's brother, is short, and has brown hair?
   a) Will
   b) Fred and Will
   c) John and Mary
   d) Will, Fred and John
   e) No answer is right.

3. Mary
   a) is short, is not John's brother, and has red hair.
   b) is tall, is not John's brother, and has brown hair.
   c) is short, is John's sister, and has brown hair.
   d) is tall, is John's sister, and has red hair.
   e) No answer is right.
4. Who had red hair?
   a) Will, Fred and John
   b) John, Mary and Will
   c) only Fred and John
   d) Mary, John and Fred
   e) No answer is right.

5. Will
   a) is short, is John's brother, and has red hair.
   b) is tall, is not John's brother, and has brown hair.
   c) is tall, is John's brother, and has brown hair.
   d) is short, is John's brother, and has red hair.
   e) No answer is right.

6. Who is tall, is not John, and has red hair?
   a) Fred and Mary
   b) Fred
   c) Will and Fred
   d) Will
   e) No answer is right.

7. John and Mary
   a) have red hair.
   b) have brown hair.
   c) are tall.
   d) are short.
   e) No answer is right.

8. Fred and Will
   a) have brown hair.
   b) are not John's brothers.
   c) are tall.
   d) have red hair.
   e) No answer is right.

II

I have five good friends. Bert lives right next door to me. Carol lives farther away than Bert. Jill lives even farther away than Carol. Dana lives farther away than Carol, but closer than Scott. I have to take a bus to his house.

1. Who lives farther away than Dana?
   a) Jill
   b) Bert
   c) Scott
   d) Carol
   e) No answer is right.
2. Carol lives
   a) farther away than Jill and Bert.
   b) farther away than Dana and Jill.
   c) closer than Scott and Bert.
   d) closer than Bert and Jill.
   e) No answer is right.

3. Who lives closer than Scott and Jill?
   a) Dana and Carol
   b) Carol and Bert
   c) only Carol
   d) only Bert
   e) No answer is right.

4. Who lives closest to me?
   a) Scott
   b) Jill
   c) Dana
   d) Bert
   e) No answer is right.

5. Who lives farthest away from me?
   a) Jill
   b) Carol
   c) Bert
   d) Scott
   e) No answer is right.

6. Jill
   a) lives farther away than Carol.
   b) lives farther away than Dana.
   c) lives closer than Carol.
   d) lives closer than Bert.
   e) No answer is right.

7. Carol
   a) lives closer than Dana.
   b) lives farther away than Jill.
   c) lives closer than Jill.
   d) lives farther away than Dana.
   e) No answer is right.

8. Dana lives farther away than
   a) Jill, Carol and Bert.
   b) Bert and Scott.
   c) Jill and Scott.
   d) Carol, Jill and Scott.
   e) No answer is right.
There are nine boys on the Bears baseball team. Mark, Bill and Ed are all taller than Henry or Dave or Bob. Jack, Ray and Al are the shortest. Henry, Jack and Mark run faster than Dave, Ray or Bill. Bob, Ed and Al are the slowest runners.

1. Who runs faster than Ray?
   a) Al and Jack
   b) Henry and Mark
   c) Mark and Dave
   d) Jack and Ed
   e) No answer is right.

2. Henry is taller than
   a) Ed.
   b) Al and Dave.
   c) Jack.
   d) Jack and Bob.
   e) No answer is right.

3. Ray runs faster than
   a) Mark and Henry.
   b) Jack, Dave and Mark.
   c) Henry and Jack.
   d) Bill.
   e) No answer is right.

4. Who is taller than Henry and faster than Ed?
   a) Dave and Bill
   b) Bill and Mark
   c) Bob and Bill
   d) Dave and Bob
   e) No answer is right.

5. Bill is taller and faster than
   a) Ed and Mark.
   b) Henry and Dave.
   c) Bob and Al.
   d) Ray and Jack.
   e) No answer is right.

6. Who is slower and taller than Dave?
   a) Jack
   b) Henry
   c) Bill
   d) Ed
   e) No answer is right.
7. Who is taller than Dave?
   a) Ed and Mark
   b) Bill and Al
   c) Mark and Al
   d) Ray and Bob
   e) No answer is right.

8. Jack is faster and shorter than
   a) Bill and Ed.
   b) Henry and Mark.
   c) Dave and Ray.
   d) Al and Bob.
   e) No answer is right.

IV

Betty, Brad, Dan, Greg and Karen go to Elm St. School. Greg is in the fifth grade, and the others are in fourth. All the fourth-graders like ice cream except Karen. Betty does not have a pet. The other fourth-graders who like ice cream do have pets. Dan has a dog.

1. Who is not in the fifth grade?
   a) only Greg.
   b) only Karen
   c) only Betty, Dan and Karen
   d) only Dan, Betty, Karen and Brad
   e) No answer is right.

2. Brad
   a) does not like ice cream.
   b) is in the fourth grade.
   c) has a pet cat.
   d) does not go to Elm St. School.
   e) No answer is right.

3. Who goes to Elm St. School?
   a) only Karen, Brad, Betty and Greg
   b) only Dan, Betty, Greg and Brad
   c) Dan, Greg, Karen, Brad and Betty
   d) everyone but Greg
   e) No answer is right.

4. Betty and Brad
   a) don't like ice cream.
   b) have pets.
   c) are not in the fourth grade.
   d) do not like pets.
   e) No answer is right.
5. Dan and Brad
   a) are in the fourth grade.
   b) do not have pets.
   c) have dogs.
   d) do not like ice cream.
   e) No answer is right.

6. Who likes ice cream?
   a) Dan, Betty and Brad
   b) Greg, Karen and Betty
   c) Brad, Betty and Karen
   d) Betty, Greg and Brad
   e) No answer is right.

7. Who is not in the fourth grade?
   a) Karen
   b) Betty
   c) Dan
   d) Greg
   e) No answer is right.

8. Betty, Dan and Brad
   a) have pets.
   b) are in the fifth grade.
   c) like ice cream.
   d) do not have pets.
   e) No answer is right.

This test was
   a) harder than most tests I've taken.
   b) easier than most tests I've taken.
   c) about the same as most tests I've taken.

APPENDIX E: The Experimental Test--Form C

DIRECTIONS: Read each story, then circle the best answer.

I

I have five good friends. Bert lives right next door to me. Carol lives farther away than Bert. Jill lives even farther away than Carol. Dana lives farther away than Carol, but closer than Scott. I have to take a bus to his house.
1. Dana lives farther away than
   a) Jill, Carol and Bert.
   b) Bert and Scott.
   c) Jill and Scott.
   d) Carol, Jill and Scott.
   e) No answer is right.

2. Jill
   a) lives farther away than Carol.
   b) lives farther away than Dana.
   c) lives closer than Carol.
   d) lives closer than Bert.
   e) No answer is right.

3. Who lives closer than Scott and Jill?
   a) Dana and Carol
   b) Carol and Bert
   c) only Carol
   d) only Bert
   e) No answer is right.

4. Carol
   a) lives closer than Dana.
   b) lives farther away than Jill.
   c) lives closer than Jill.
   d) lives farther away than Dana.
   e) No answer is right.

5. Who lives closest to me?
   a) Scott
   b) Jill
   c) Dana
   d) Bert
   e) No answer is right.

6. Who lives farthest away from me?
   a) Jill
   b) Carol
   c) Bert
   d) Scott
   e) No answer is right.

7. Who lives farther away than Dana?
   a) Jill
   b) Bert
   c) Scott
   d) Carol
   e) No answer is right.
8. Carol lives
   a) farther away than Jill and Bert.
   b) farther away than Dana and Jill.
   c) closer than Scott and Bert.
   d) closer than Bert and Jill.
   e) No answer is right.

II

Betty, Brad, Dan, Greg and Karen to Elm St. School. Greg is in the fifth grade, and the others are in fourth. All the fourth-graders like ice cream except Karen. Betty does not have a pet. All the other fourth-graders who like ice cream do have pets. Dan has a dog.

1. Who goes to Elm St. School?
   a) only Karen, Brad, Betty and Greg
   b) only Dan, Betty, Greg and Brad
   c) Dan, Greg, Karen, Brad and Betty
   d) everyone but Greg
   e) No answer is right.

2. Who likes ice cream?
   a) Dan, Betty and Brad
   b) Greg, Karen and Betty
   c) Brad, Betty and Karen
   d) Betty, Greg and Brad
   e) No answer is right.

3. Who is not in the fourth grade?
   a) Karen
   b) Betty
   c) Dan
   d) Greg
   e) No answer is right.

4. Who is not in the fifth grade?
   a) only Greg
   b) only Karen
   c) only Betty, Dan and Karen
   d) only Dan, Betty, Karen and Brad
   e) No answer is right.

5. Brad
   a) does not like ice cream.
   b) is in the fourth grade.
   c) has a pet cat.
   d) does not go to Elm St. School.
   e) No answer is right.
6. Betty and Brad
   a) don't like ice cream.
   b) have pets.
   c) are not in the fourth grade.
   d) do not have pets.
   e) No answer is right.

7. Betty, Dan and Brad
   a) have pets.
   b) are in the fifth grade.
   c) like ice cream.
   d) do not have pets.
   e) No answer is right.

8. Dan and Brad
   a) are in the fourth grade.
   b) do not have pets.
   d) have dogs.
   d) do not like ice cream.
   e) No answer is right.

III

John had two brothers who were both tall. Their names were Will and Fred. John's sister, who was short, was named Mary. John liked Fred better than either of the others. All these children except Will had red hair. He had brown hair.

1. Mary
   a) is short, is not John's brother, and has red hair.
   b) is tall, is not John's brother, and has brown hair.
   c) is short, is John's sister, and has red hair.
   d) is tall, is John's sister, and has brown hair.
   e) No answer is right.

2. Fred and Will
   a) have brown hair.
   b) are not John's brothers.
   c) are tall.
   d) have red hair.
   e) No answer is right.

3. Will
   a) is short, is John's brother, and has red hair.
   b) is tall, is not John's brother, and has brown hair.
   c) is short, is not John's brother, and has red hair.
   d) is tall, is John's brother, and has brown hair.
   e) No answer is right.
4. Who had red hair?
   a) Will, Fred and John
   b) John, Mary and Will
   c) only Fred and John
   d) Mary, John and Fred
   e) No answer is right.

5. Who is John's brother, is short, and has brown hair?
   a) Will
   b) Fred and Will
   c) John and Mary
   d) Will, Fred and John
   e) No answer is right.

6. Which children are not John's brother?
   a) Will and Mary
   b) Fred and John
   c) John and Mary
   d) Will and Fred
   e) No answer is right.

7. John and Mary
   a) have red hair.
   b) have brown hair.
   c) are tall.
   d) are short.
   e) No answer is right.

8. Who is tall, is not John, and has red hair?
   a) Fred and Mary
   b) Fred
   c) Will and Fred
   d) Will
   e) No answer is right.

IV

There are nine boys on the Bears baseball team. Mark, Bill and Ed are all taller than Henry or Dave or Bob. Jack, Ray and Al are the shortest. Henry, Jack and Mark run faster than Dave, Ray or Bill. Bob, Ed and Al are the slowest runners.

1. Ray runs faster than
   a) Mark and Henry.
   b) Jack, Dave and Mark.
   c) Henry and Jack.
   d) Bill.
   e) No answer is right.
2. Who runs faster than Ray?
   a) Al and Jack
   b) Henry and Mark
   c) Mark and Dave
   d) Jack and Ed
   e) No answer is right.

3. Jack is faster and shorter than
   a) Bill and Ed.
   b) Henry and Mark.
   c) Dave and Ray.
   d) Al and Bob.
   e) No answer is right.

4. Who is taller than Henry and faster than Ed?
   a) Dave and Bill
   b) Bill and Mark
   c) Bob and Bill
   d) Dave and Bob
   e) No answer is right.

5. Bill is taller and faster than
   a) Ed and Mark.
   b) Henry and Dave.
   c) Bob and Al.
   d) Ray and Jack.
   e) No answer is right.

6. Henry is taller than
   a) Ed.
   b) Al and Dave.
   c) Jack.
   d) Jack and Bob.
   e) No answer is right.

7. Who is taller than Dave?
   a) Ed and Mark
   b) Bill and Al
   c) Mark and Al
   d) Ray and Bob
   e) No answer is right.

8. Who is slower and taller than Dave?
   a) Jack
   b) Henry
   c) Bill
   d) Ed
   e) No answer is right.
This test was
a) harder than most tests I've taken.
b) easier than most tests I've taken.
c) about the same as most tests I've taken.