To the Graduate Council:

I am submitting herewith a dissertation written by Maurice Houston Field entitled "Some Effects of Elementary Science Textbook Readability Levels on Science Achievement of Elementary Students with Low, Average, and High Reading Abilities." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Education, with a major in Curriculum and Instruction.

W. W. Wyatt, Major Professor

We have read this dissertation
and recommend its acceptance:

A Paul Wishart

Joel W Morris

Paul C Burns

Accepted for the Council:

L Evans Roth
Vice Chancellor
Graduate Studies and Research

SOME EFFECTS OF ELEMENTARY SCIENCE TEXTBOOK READABILITY LEVELS ON SCIENCE ACHIEVEMENT OF ELEMENTARY STUDENTS WITH LOW, AVERAGE, AND HIGH READING ABILITIES

A Dissertation

Presented for the

Doctor of Education

Degree

The University of Tennessee, Knoxville

Maurice Houston Field

March 1982

3058031

ABSTRACT

The purpose of this study was to investigate some of the effects of elementary science textbook readability levels on science achievement of elementary school students with low, average, and high reading abilities as measured by the Metropolitan Achievement Tests: Science.

The total reading achievement score obtained on the Metropolitan Achievement Tests: Elementary Form F was used to subdivide 111 students into three reading ability groups. The three groups were found to be significantly different when compared on total reading achievement scores and later were found to be significantly different when compared on science achievement scores obtained from the Metropolitan Achievement Tests: Intermediate Form G administered in April 1980. The gains in total reading achievement were not significantly different from gains in science achievement for any of the three reading ability groups. Gains were calculated from scores on the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

Investigation of the relationship between ability to read the Metropolitan Achievement Tests: Intermediate Form G and scores above 6.7 on science achievement indicated that students who scored above 6.0 on total reading achievement tended to score above 6.7 on science achievement. Study of the relationship between ability to read the sixth grade science textbook and scores above 6.7 on science achievement indicated that students who scored above 7.0 on total reading achievement

tended to score above 6.7 on science achievement. Students who scored below 7.0 on total reading achievement had an equal opportunity of scoring above or below 6.7 on science achievement.

A review of the findings led to the following conclusions:

1. Science achievement as measured by the Metropolitan Achievement Tests was affected by total reading achievement.

2. Gains in science achievement scores and total reading achievement scores tended to progress concurrently.

3. Science achievement scores were adversely affected when students responded to an achievement test with an estimated readability level that exceeded the total reading achievement level of the individual student.

4. Science achievement scores were adversely affected when students studied a basal elementary school textbook which had an estimated readability level which exceeded the total reading achievement level of the individual student.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	THE PROBLEM	1
	Purpose of the Study	1
	Need for the Study	1
	Overview of Procedures	3
	Hypotheses	5
	Assumptions	7
	Limitations	8
	Definitions of Terms	8
	Organization of Chapters	9
II.	REVIEW OF THE LITERATURE	10
	Readability defined	16
	Evolution of readability	17
	Readability and elementary school science textbooks	29
	Readability and secondary school science textbooks	47
	Comparison of readability formulas	55
III.	PROCEDURES	67
	Selection of Elementary School Science Textbook Series	67
	Readability formula selection	67
	Criteria for textbook series selection	70
	Selection of Achievement Tests	72
	Selection of Reading Ability Groups	73
	Readability of Achievement Tests	76
	Selection of Statistical Procedures	82
	Mann-Whitney U test	84
	Wilcoxon Matched-Pairs Signed-Ranks test	86
	2 × 2 contingency table	88
	Fisher's Exact Probability test	89
IV.	FINDINGS	91
	Analysis of Fourth Grade Total Reading Achievement	92
	Analysis of Sixth Grade Science Achievement	94
	Analysis of Total Reading Achievement and Science Achievement Gains	95
	Analysis of the Effects of Readability Estimates on Science Achievement	98

CHAPTER PAGE

Analysis of Data by Inspection of Individual Reading Ability Groups and the Total Reading Ability Group 105

V. SUMMARY AND CONCLUSIONS 114

Summary 114
Conclusions 118
Implications 121

BIBLIOGRAPHY 123

APPENDIX 132

VITA 134

LIST OF TABLES

TABLE		PAGE
I.	Estimated Readability Levels of Fifth and Sixth Grade Elementary School Science Textbooks Adopted in Tennessee for September 1975 through June 1980	71
II.	Correlations Produced by Comparing 1977-78 Reading Achievement Subpart Scores and Science Achievement Scores for 111 Students in 1978-79 and 1979-80	75
III.	Metropolitan Achievement Tests Scores in Grade Level Equivalents for Total Reading Achievement and Science Achievement Compiled for a Low Reading Ability Group during the Period 1977-80	78
IV.	Metropolitan Achievement Tests Scores in Grade Level Equivalents for Total Reading Achievement and Science Achievement Compiled for an Average Reading Ability Group during the Period 1977-80	79
V.	Metropolitan Achievement Tests Scores in Grade Level Equivalents for Total Reading Achievement and Science Achievement Compiled for a High Reading Ability Group during the Period 1977-80	81
VI.	Mann-Whitney U Test Analysis of Fourth Grade Total Reading Achievement Scores of Reading Ability Groups	94
VII.	Mann-Whitney U Test Analysis of Sixth Grade Science Achievement Scores of Reading Ability Groups	96
VIII.	Wilcoxon Matched-Pairs Signed-Ranks Test Analysis of Total Reading Achievement Gains and Science Achievement Gains of Reading Ability Groups	99
IX.	Fisher's Exact Probability Test Analysis of the Effects of Readability Estimates for the Metropolitan Achievement Tests: Science and the Sixth Grade Science Textbook on Grade Level Science Achievement	105
X.	Individual Analysis of Data by Group for Comparisons of Total Reading Achievement to 6.0 and 7.0 with Science Achievement Scores Obtained from the Metropolitan Achievement Tests: Intermediate Form G	110

LIST OF FIGURES

FIGURE		PAGE
1.	Fry Readability Graph	69
2.	Selection of Low, Average, and High Reading Ability Groups	77
3.	Presentation of the Effect of Total Reading Achievement on Grade Level Science Achievement for a Low Reading Ability Group	106
4.	Presentation of the Effect of Total Reading Achievement on Grade Level Science Achievement for an Average Reading Ability Group	107
5.	Presentation of the Effect of Total Reading Achievement on Grade Level Science Achievement for a High Reading Ability Group	108
6.	Presentation of the Combined Reading Ability Groups	113

CHAPTER I

THE PROBLEM

I. PURPOSE OF THE STUDY

The purpose of this study was to investigate some of the effects of elementary science textbook readability levels on science achievement of elementary school students with low, average, and high reading abilities as measured by the Metropolitan Achievement Tests: Science.

II. NEED FOR THE STUDY

Reports of declining test scores have caused a considerable amount of concern which has been publicly discussed and written about frequently in the media.[1, 2] Harnishfeger and Wiley reported that most standardized test scores steadily increased through the 1940's, 50's and into the middle 60's before starting to decline.[3] Furthermore, Harnishfeger and Wiley reported that data collected on children through the fourth grade level showed no decline as compared to previous years. However, there

---

[1]J. Stanley Ahmann et al., "Science Achievement: The Trend is Down," The Science Teacher, 42 (September 1975), 23.

[2]Richard L. Ferguson, "ACT Scores—What's behind the Decline?" NASSP Bulletin, 60 (May 1976), 71.

[3]Annegret Harnishfeger and David E. Wiley, "The Marrow of Achievement Test Score Declines," Educational Technology, 16 (June 1976), 5-14.

was the beginning of a decline in the fifth grade which continued through the twelfth grade level.[4]

Science taught at the fifth and sixth grade levels was generally taught by a generalist and was often dependent upon the reading of the textbook. It would appear that one of the possible limiting factors in science achievement may have been the student's inability to read the adopted textbook.

Field et al. reported the approximate readability levels for the twelve elementary school science textbook series adopted in Tennessee in 1975. Their findings indicated that at least eleven of the twelve series were written at grade level or below for grades one through three. Textbooks written for the fourth grade were written above grade level in seven series. Fifth grade textbooks were above grade level in ten series, while sixth grade textbooks were written above grade level in five series.[5]

Investigation of the effects of these data on science achievement appears pertinent to the case of declining test scores. Thus the readability level of the adopted textbook which the student studied may have been a factor in the cause of the decline in science achievement

---

[4]Annegret Harnishfeger and David E. Wiley, "Achievement Test Scores Drop, So What?" Educational Researcher, 5 (March 1976), 5-12.

[5]Maurice Houston Field et al., "The Approximate Reading Level Ranges and Means by Grade Level for Twelve Elementary School Science Textbook Series," Science Education Bulletin 101 (Martin, Tennessee: Department of Elementary and Secondary Education, The University of Tennessee at Martin, 1975), pp. 1-34.

throughout the public school years. Of special value is the possible relationship between the increase in readability level at the fourth grade level and the data cited by Harnishfeger and Wiley describing the beginning of test score declines after the fourth grade level.

III. OVERVIEW OF PROCEDURES

This study involved 111 students selected from a population of 153 sixth grade students. These students had been annually tested with the Metropolitan Achievement Tests and had studied science from a basal science textbook series which included textbooks that possessed estimated readabilities above the grade level for which they were intended.

The 111 students selected for this study had been tested at the fourth grade level with the Metropolitan Achievement Tests: Elementary Form F[6] which was administered in October 1977. They had also been tested at the fifth grade level with the Metropolitan Achievement Tests: Intermediate Form F[7] in October 1978, and sixth grade level with the Metropolitan Achievement Tests: Intermediate Form G[8] in April 1980.

The total reading achievement score obtained on the Metropolitan Achievement Tests: Elementary Form F was used to subdivide the 111

---

[6]Walter N. Durost et al., Metropolitan Achievement Tests: Elementary Form F (New York: Harcourt Brace Jovanovich, 1972).

[7]Walter N. Durost et al., Metropolitan Achievement Tests: Intermediate Form F (New York: Harcourt Brace Jovanovich, 1970).

[8]Walter N. Durost et al., Metropolitan Achievement Tests: Intermediate Form G (New York: Harcourt Brace Jovanovich, 1970).

students into three reading ability groups. The low reading ability group was composed of those students with total reading achievement scores of 4.0 or less. The average reading ability group had total reading achievement scores of 4.1 to 6.0. The high reading ability group possessed total reading achievement scores of 6.1 and above. The three groups contained 27, 50, and 34 members respectively.

In order to establish that the reading ability groups were significantly different when compared in total reading achievement, the Mann-Whitney U test was employed. A one-tailed region of rejection was used with a .05 level of significance.

Science achievement data obtained from the Metropolitan Achievement Tests: Intermediate Form G administered in April 1980 were compared by the Mann-Whitney U test. A two-tailed region of rejection was selected with a .05 level of significance.

Total reading achievement scores and science achievement scores were taken from the Metropolitan Achievement Tests: Intermediate Forms F and G which had been administered respectively in October 1978 and April 1980. Gains in total reading achievement and science achievement were computed. These gains were then subjected to test by the Wilcoxon Matched-Pairs Signed-Ranks test. A two-tailed region of rejection with a .05 level of significance was selected.

In order to determine the relation between readability of materials and the attainment of grade level science achievement the Fisher's Exact Probability test was selected. Attempts were made to use both the 2×2 contingency table and the 2×2 contingency table with the Yate's

correction for continuity. Since neither of these were appropriate due to small cell entries, the Fisher's Exact Probability test was selected. The tests were dependent on making comparisons of total reading achievement with the estimated readabilities of the science portion of the Metropolitan Achievement Tests: Intermediate Form G and the sixth grade elementary school science textbook. The estimated readabilities were 6.0 for the former and 7.0 for the latter. The aforementioned reading comparisons were made in connection with the attainment of grade level science achievement which was deemed to be 6.7. The region of rejection was two-tailed with the level of significance set at .05.

IV. HYPOTHESES

1. There is no significant difference between a low reading ability group and an average reading ability group as compared on total reading achievement scores measured by the Metropolitan Achievement Tests: Elementary Form F administered in October 1977.

2. There is no significant difference between an average reading ability group and a high reading ability group as compared on total reading achievement scores measured by the Metropolitan Achievement Tests: Elementary Form F administered in October 1977.

3. There is no significant difference between a low reading ability group and an average reading ability group as compared on science achievement scores measured by the Metropolitan Achievement Tests: Intermediate Form G administered in April 1980.

4. There is no significant difference between an average reading ability group and a high reading ability group as compared on science

achievement scores measured by the Metropolitan Achievement Tests: Intermediate Form G administered in April 1980.

5. There is no significant difference between total reading achievement gains and science achievement gains for a low reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

6. There is no significant difference between total reading achievement gains and science achievement gains for an average reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

7. There is no significant difference between total reading achievement gains and science achievement gains for a high reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

8. There is no significant difference between low reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

9. There is no significant difference between average reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

10. There is no significant difference between high reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

11. There is no significant difference between low reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

12. There is no significant difference between average reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

13. There is no significant difference between high reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

V. ASSUMPTIONS

1. Achievement in science at the fifth grade and sixth grade levels is dependent upon the student's reading the adopted textbook.

2. Students not demonstrating a reading ability equal to or higher than the estimated readability level of the elementary science textbook studies will perform poorly on an achievement test in science.

3. The readability level of a science textbook or achievement test can be estimated by such procedures as the Dale-Chall, Flesch, Fry Readability Graph, Gunning-Fog, and Spache.

4. The reading ability of the students can be estimated by the use of standardized tests such as the California, Gates-MacGinite, Iowa Silent, Metropolitan, Sequential Test of Educational Progress and Stanford.

5. The Metropolitan Achievement Tests: Science is written on a readability level which does not serve as a limiting factor in obtaining science achievement scores for the majority of students.

VI. LIMITATIONS

This study was limited to sixth grade students selected from a school system which adopted an elementary school science textbook series which contained fifth and sixth grade textbooks with readability estimates above the grade level for which they were intended as determined by the Fry Readability Graph. In each instance the students had studied science textbooks which were part of an elementary school science textbook series adopted when the students were in the fourth grade.

VII. DEFINITIONS OF TERMS

Readability refers to the reading ability necessary for a reader to read a particular passage.[9] Some of the principal bases on which readability is determined are number of words in a sample selection, average length of sentences, and number of syllables in a word.

---

[9]Russell G. Stauffer, Directing the Reading-Thinking Process (New York: Harper and Row, 1969), p. 363.

Reading ability is the reading achievement of a student stated in terms of curricular validity.[10] Reading ability is normally determined by standardized tests which can be converted to grade level scores as well as percentiles and stanines which would show how the student compares with a group on which the tests were standardized.

Science achievement as measured by the Metropolitan Achievement Tests: Science emphasizes the ability to recall important but specific information in the area of science.[11]

VIII. ORGANIZATION OF CHAPTERS

Purpose of the study, the need for the study, overview of procedures, assumptions, the limitations, definition of terms, and the organization of chapters have been included in Chapter I.

The review of the literature is confirmed in Chapter II.

The research procedures are outlined and discussed in Chapter III.

The findings of the study are reported in Chapter IV.

The summary and conclusions are reviewed in Chapter V.

---

[10]Edward Fry, Reading Instruction for Classroom and Clinic (New York: McGraw-Hill Book Company, 1972), p. 25.

[11]Gilbert Sax, Principles of Educational Measurement and Evaluation (Belmont, California: Wadsworth Publishing Company, 1974), p. 387.

CHAPTER II

REVIEW OF THE LITERATURE

The most authoritative data on science achievement have been produced as a result of a survey reported as the National Assessment of Educational Progress. This report indicated that 9-year olds, 13-year-olds, and 17-year-olds, tended to perform less well in 1972-73 than they had in 1969-70.[1]

The Tennessee Assessment of Educational Progress substantiated, in part, the research cited from the National Assessment of Science, 1969-73. In April 1975, the Tennessee State Testing and Evaluation Center began a study to determine trends in Tennessee. The study was based on the Metropolitan Achievement Tests scores which had been obtained as a result of a large number of schools consistently administering the examinations at the fifth grade level. In comparing data taken from the years 1970 to 1974 in all subject matter areas except word knowledge, declines were found. In the area of science achievement there was a drop of 1.9 months in the grade level equivalent.[2]

---

[1]National Assessment of Educational Progress, National Assessments of Science, 1979 and 1973: A Capsule Description of Changes in Science Achievement (Science Report No. 04-5-00. Washington: U.S. Government Printing Office, 1975), p. 13.

[2]Tennessee Department of Education, Tennessee Assessment of Educational Progress Results Report: Reading, 1974-75 (Nashville, Tennessee: State Department of Education, 1975), p. 76.

The National Assessment of Educational Progress centered on science again in 1976-77. Highlights of this study were:

> 1. The downward trend in science achievement observed from the first to the second assessment appears to be lessening for 9- and 13-year-olds in the third assessment. For these two age populations, it appears that a continued decline in achievement in physical science is accompanied by some stability in achievement in biology.
> 2. The achievement level of 17-year-olds continues to decline. Seventeen-year-olds declined in performance from the first to the second assessment and from the second to the third assessment.[3]

An indepth study of the report indicated that 9-year-olds compiled a mean change of -.1, while 13-year-olds showed a mean change of -.7. Mean change was defined as mean change in percentage correct from 1969-70 data. Although neither change was significant, the 13-year-olds continued to decline. However, the decline was less than the mean change of -1.7 found between 1969-70 and 1972-73.[4]

A number of ideas have been expressed as to why the declines in test scores have taken place. There is at this time insufficient evidence to substantiate probable cause.

Nelson cited some factors he felt to be pertinent to the general decline in test scores. Some of these are: lack of discipline in the schools, breakdown in the family unit, lack of proper teacher training,

---

[3]National Assessment of Educational Progress, Three National Assessments of Science: Changes in Achievement, 1969-77 (Science Report No. 08-5-00) Denver: Education Commission of the States, 1978), p. 13.

[4]Ibid., p. 6.

less effective teaching of skill subjects such as reading, decline in teaching of factual material, lack of desire for education, insufficient time allowed for teaching science, particularly at the elementary level, and increase in the school's holding power.[5]

Sapone and Giuliano also proposed a number of possibilities for the general decline. Some of these are recent curricular changes made in response to social needs, massive reorganization of school systems, lower average number of days attended by students, current teaching force is better trained but less experienced, increased television ownership, the amount of time spent watching television, and the increase in illegitimate births between 1948 and 1974.[6]

Some views have been proposed which are specifically concerned with the declines in science achievement. One of these views concerned the phenomenon surrounding Sputnik. According to this theory, science achievement increases were produced by a strong emphasis on science while the present declines were brought about by a return to a more normal atmosphere.[7]

Funk proposed that students have learned less science or possibly that tests used to measure science achievement have been at fault.[8]

---

[5]Donald J. Nelson, "What's behind the Test Score Decline?" NASSP Bulletin, 60 (May 1976), 78-79.

[6]Carmelo V. Sapone and Joseph R. Giuliano, "The Test Score Decline: Are the Public Schools the Scapegoat?" Educational Technology, 16 (June 1976), 43-44.

[7]National Assessment of Educational Progress, loc. cit.

[8]H. James Funk, "What Research Says to the Teacher: An Examination of the Levels of Questions on Standardized Tests for Elementary Science," Science and Children, 14 (January 1977), 30.

Thelen wrote concerning the continual reporting of science teachers that students cannot read the textbook.[9]

In 1969 Ramsey and Howe made an analysis of research related to the instructional procedures of the elementary school. Their review covered a number of subjects which include programmed instruction, individualized instruction, ability grouping, and use of reading materials. With regard to the use of reading materials Ramsey and Howe concluded:

> Research in the area of the use of reading materials is indeed thin. More and more, textbooks and other materials directed to the elementary pupil are coming onto the market, yet the role of reading materials in science instruction has had little recent evaluation.[10]

In 1972 Mallinson conducted a review of research limited to reading in the sciences in which he considered both elementary and secondary school science. Mallinson singled out television and inquiry oriented programs as having affected the production of research related to reading in the sciences. He was forced to conclude:

> 1. A search of the literature and a perusal of the titles of papers presented at professional meetings indicate a great decline in the past decade in the numbers of research investigations dealing with problems of reading in science.
>
> 2. From the few studies undertaken in recent years it is difficult to ascertain any direction. Few research

---

[9]Judith N. Thelen, "Using the Cloze Test with Science Textbooks," Science and Children, 12 (November-December 1974), 26.

[10]George A. Ramsey and Robert W. Howe, "An Analysis of Research Related to Industrial Procedures in Elementary School Science" Science and Children, 6 (April 1969), p. 31.

> investigations appear to be replications, extensions, or corollaries of other studies of the period.[11]

In addition Mallinson cautioned that even though many science programs have moved toward "inquiry" programs from "read about" programs, students must still be able to read the "foci" for inquiry and be able to understand the task that is to be performed. Mallinson summarized:

> . . . although several factors have apparently resulted in a temporary dearth of research studies related to problems of reading in science, they may well have opened many new avenues in the future for productive and meaningful studies.[12]

Learning in science has been shown to be highly related to reading comprehension according to Blooms's observations of the International Association for the Evaluation of Educational Achievement studies.[13] Proceeding a step further, Piltz and Sund wrote that "reading is an important tool in learning science, and textbooks may insure a developmental program and serve a unique function in the learning process."[14]

According to Bond and Tinker, reading in the content areas must be attuned to individual differences and cannot omit the value of suitable reading materials. In order for a child to read, there must

---

[11]George G. Mallinson, "Reading in the Sciences: A Review of the Research," Reading for the Content Areas (New York: International Reading Association, 1972), pp. 146-147.

[12]Ibid., p. 148.

[13]Benjamin S. Bloom, "Implications of the IEA Studies for Curriculum and Instruction," Educational Policy and International Assessment: Implications of the IEA Surveys of Achievement (Berkeley, California: McCutchan Publishing Corporation, 1975), p. 72.

[14]Albert Piltz and Robert Sund, Creative Teaching of Science in the Elementary School (Boston: Allyn and Bacon, 1970), p. 107.

be materials available that the child is capable of reading.[15] Pooley extended this theme by suggesting that the right books must be available at the appropriate time in a child's reading development to make him want to continue to read.[16]

Mills and Richardson in 1963 studied the grade levels of texts supplied by publishers. After grading more than 200 books recommended for use by elementary children, they discovered considerable inconsistency between publishers' estimates and estimates of readability as determined by the Spache and Dale-Chall formulas. An important conclusion of the investigation was that many books were assigned grade-level placements by authors and consultants based on personal judgement rather than by a more consistent method.[17] In 1972 Dale reported that "too many textbooks, pamphlets, encyclopedias . . . are written at a grade level above that of the readers for whom they were prepared."[18]

Carin and Sund reported that "one of the most glaring limitations of science textbooks in the elementary school is the problem of readability."[19] Mallinson in 1955 found that science textbooks were too

---

[15]Guy L. Bond and Miles A. Tinker, Reading Difficulties, Their Diagnosis and Correction (New York: Appleton-Century-Crofts, 1957), pp. 52-53.

[16]Robert C. Pooley, "Reading and the Language Arts," Development in and through Reading, The Sixteenth Yearbook of the National Society for the Study of Education, Part I, 1961, pp. 52-53.

[17]R. E. Mills and Jean R. Richardson, "What Do Publishers Mean by 'Grade Level'?" The Reading Teacher, 16 (March 1963), 359-362.

[18]Edgar Dale, Building a Learning Environment (Bloomington, Indiana: Phi Delta Kappa, Inc., 1972), p. 30.

[19]Arthur A. Carin and Robert B. Sund, Teaching Modern Science (Columbus, Ohio: Charles E. Merrill Publishing Company, 1970), p. 143.

difficult for the children for whom they were written.[20] In 1965 Ottley reached the same conclusion based on his study of fourth, fifth, and sixth grade textbooks.[21] In 1975 Field et al. published similar findings for the fourth, fifth, and sixth grade textbooks evaluated.[22]

Readability Defined

Readability has been defined by Good as "quality of a piece of reading matter that makes it readable and understandable."[23] Stauffer defined readability as the "reading ability required by a reader to grasp a selection."[24] McLaughlin in 1968 defined readability as "the degree to which a given class of people find certain reading matter compelling and necessarily comprehensible."[25]

Dale and Chall in 1948 defined readability more fully:

---

[20]George Greisen Mallinson, Harold E. Sturm, and Lois Marion Mallinson, "The Reading Difficulty of Unit-Type Textbooks for Elementary Science," Science Education, 39 (December 1955), 410.

[21]LeRoy Ottley, "Readability of Science Textbooks for Grades Four, Five, Six," School Science and Mathematics, 65 (April 1965), 363-366.

[22]Maurice Houston Field et al., "The Approximate Reading Level Ranges and Means by Grade Level for Twelve Elementary School Science Textbook Series," Science Education Bulletin 101 (Martin, Tennessee: Department of Elementary and Secondary Education, The University of Tennessee at Martin, 1975), pp. 1-34.

[23]Carter V. Good, Dictionary of Education. Second Edition (New York: McGraw-Hill Book Company, 1959), p. 442.

[24]Russell G. Stauffer, Directing the Reading-Thinking Process (New York: Harper and Row, 1969), p. 363.

[25]G. H. McLaughlin, "SMOG Grading—A New Readability Formula," Journal of Reading, 12 (May 1969), 639-646.

> Readability is the sum total (including interactions) of all those elements within a given piece of material that affect the success that a group of readers have with it. The success is the extent to which they understand it, read at optimum speed, and find it interesting.[26]

Evolution of Readability

In 1921 an early attempt at measuring readability was made with the publications by Thorndike of a comprehensive count of word frequencies which listed 10,000 words rated by frequency of occurrence and subdivided into groups of 1,000 words.[27] He followed this original work with two other efforts completed in 1931 and 1944 respectively which involved lists of 20,000 and 30,000 words.[28,29]

Lively and Pressey made one of the earlier attempts at developing a readability formula as they explored a method of estimating vocabulary difficulty in junior high school science texts in 1923. Their work was based on Thorndike's List of Ten Thousand Words and suggested that four possible indexes of vocabulary burden might separate hard from easy words. The indexes included: (1) number of different words in a passage; (2) difficulty assigned by Thorndike's word list; (3) count of words not

---

[26]Edgar Dale and Jeanne L. Chall, "The Concept of Readability," Elementary English, 26 (January 1949), 23.

[27]Edward L. Thorndike, A Teacher's Word Book (New York: Teachers College Bureau of Publications, Columbia University, 1921).

[28]Edward L. Thorndike, A Teacher's Word Book of 20,000 Words (New York: Teachers College Bureau of Publications, Columbia University, 1931).

[29]Edward L. Thorndike and Irving Lorge, The Teacher's Word Book of 30,000 Words, (New York: Teachers College, Bureau of Publications, Columbia University, 1944).

on Thorndike's list; and (4) a combination of (2) and (3) referred to as a weighted median index. The results showed that only the weighted median index gave a reasonable estimate of difficulty.[30]

Vogel and Washburne followed with another procedure based on Thorndike's list in an effort to establish the readability of children's books in 1928. The variables which were considered in the formula included the number of different words that occurred in 1,000 word samples, the number of prepositions occurring in each sample, the number of uncommon words in each sample, and the number of simple sentences found in a sample of 75 sentences. The words considered to be uncommon were those not included on Thorndike's list of 1,500 most common words. They let x_2 represent the number of different words in a 1,000 word sample; x_3, the number of prepositions; x_4, the number of uncommon words; and x_5, the number of simple sentences in 75. The resulting formula was:

$$x_1 \text{ or Reading Score} = .085x_2 + .101x_3 + .064x_4 - .411x_5 + 17.43.$$

The reading score which resulted was considered equivalent to the same score on the paragraph meaning portion of the Stanford Achievement Test. This score was then translated to a grade level by means of a table.[31]

Lewerenz in 1929 reported an investigation of the relationship between letters of the alphabet and readability which used Webster's

---

[30]Bertha Lively and S. L. Pressey, "A Method for Measuring the Vocabulary Burden of Textbooks," Educational Administration and Supervision, 9 (October 1923), 389-398.

[31]Mabel Vogel and Carleton Washburne, "An Objective Method of Determining Grade Placement of Children's Reading Material," Elementary School Journal, 27 (January 1928), 373-381.

Elementary Dictionary as the standard. The findings of the study indicated that in easy reading material the words beginning with the letters w, h, and b occurred more frequently, while in difficult reading material the letters i and e occurred more often.[32]

Johnson was one of the first to study the effects of polysyllabic words on readability. In his 1930 study he concluded that tests written in monosyllabic words were less difficult than those which included large numbers of polysyllabic words.[33]

Dale and Tyler in 1934 attempted to devise a readability formula that would identify the difficulty level of materials written for adults of limited reading abilities. The difficulty level of the materials was determined by using the average reading comprehension scores of adults with limited reading abilities as assessed by tests which were constructed on selected criteria. The criteria consisted of 74 selections from newspapers, magazines, textbooks, selections adapted from children's health books, and very simple selections written by the authors. The results indicated that three factors were significant in predicting the difficulty of reading material. These were: (1) number of different technical words; (2) number of hard nontechnical words; and (3) number

---

[32]Alfred A. Lewerenz, "Measurements of Difficulty of Reading Materials," Educational Research Bulletin, 8 (January 1929), 11-16.

[33]George R. Johnson, "An Objective Method of Determining Readability," Journal of Educational Research, 21 (April 1930), 283-287.

of indeterminate clauses. A correlation of .51 with the previously cited criteria was found.[34]

Gray and Leary in 1935 developed a readability formula that divided reading materials into five levels of difficulty. The investigators chose 48 selections graded in difficulty by comprehension scores of readers. These criteria were used to determine the relative importance of 289 factors related to reading. In the final analysis five factors emerged as significant and were set into a regression equation. The five significant factors were: (1) number of personal pronouns; (2) number of different hard words; (3) average number of words per sentence; (4) percentage of different words; and (5) number of prepositional phrases. A correlation of .64 with the 48 selections was found.[35]

Lorge in 1939 used the five factors identified by Gray and Leary to which he added a weighted index of word difficulty based on the level of the word according to Thorndike's List of Ten Thousand Words. His criteria were studies previously standardized and graded according to level of difficulty which were taken from McCall-Crabbs Standard Test Lessons in Reading. The best predictors of reading difficulty were: (1) number of different hard words not on the Dale List of 769 Words; (2) average sentence length; and (3) number of prepositional phrases. His prediction formula thus consisted of a regression equation involving

---

[34]Edgar Dale and Ralph Tyler, "A Study of the Factors Influencing the Difficulty of Reading Materials for Adults of Limited Reading Ability," Library Quarterly, 4 (July 1934), 384-412.

[35]William Gray and Bernice Leary, What Makes a Book Readable? (Chicago: University of Chicago Press, 1935).

these three factors where x_2 represented the average sentence length, x_3 represented the number of prepositional phrases per 100 words, and x_4 represented the number of different hard words not on the Dale List of 769 Words. The formula was:

x_1 or grade placement = $.07x_2 + .1301x_3 + .1073x_4 + 1.6126$. A correlation of .77 was found between his formula and selected criteria.[36]

In 1939 Yoakum designed a formula which was similar to Lorge's formula. His major purpose was to simplify the calculation of readability by a formula. The work was not published until 1955 and was based on a weighted index of vocabulary difficulty obtained by assigning the index number from Thorndike's List of Twenty Thousand Words in a reading sample which fell into the fourth thousand and above frequency. The index numbers were totaled on the passage selected. The index total was divided by the number of pages to determine a page index average. The grade level of the passage was determined by referring to The Yoakum Difficulty Scale.[37]

Flesch in 1943 developed a readability formula which considered two variables that had been included in previous formulae. He examined the number of morphemes and the number of abstract words contained in a given passage as well as the standard variables of difficult words and

---

[36]Irving Lorge, "Predicting Reading Difficulty of Selections for Children," Elementary English Review, 16 (October 1939), 229-233.

[37]Gerald A. Yoakum, Basal Reading Instruction (New York: McGraw-Hill Book Company, 1955), pp. 329-337.

sentence length. His study concentrated on the McCall-Crabbs Standard Test Lessons in Reading and materials written for adult readers which had been ranked in order of difficulty. Three factors emerged from the study which proved to be of significance. These were: (1) average sentence length; (2) number of affixed morphemes; and (3) number of personal references. Flesch's formula was:

Grade Placement = $.1338x_s + .0645x_m - .0659x_h - .7502$,

where x_s represented average sentence length, x_m represented the number of affixed morphemes, and x_h represented the number of personal references. In the formula grade placement referred to a 50 percent comprehension level, a 75 percent level of comprehension was estimated by adding the constant 4.2498 to the grade placement.[38,39]

After Flesch published his work, Lorge introduced another formula for measuring difficulty. This effort was reported in order to correct an error that had been discovered in the original formula. The new formula again required x_2 to represent average sentence length in words, x_3 to represent the number of prepositional phrases per 100 words, and x_4 to represent the number of different hard words not on the Dale List of 769 Words. The formula was:

x_1 or grade placement = $.06x_2 + .10x_3 + .10x_4 + 1.99$.

---

[38]Rudolf Flesch, Marks of Readable Style (New York: Bureau of Publications, Teachers College, Columbia University, 1940).

[39]Rudolf Flesch, "Estimating Comprehension of Difficulty of Magazine Articles," Journal of General Psychology, 27 (January 1943), 63-80.

This formula allowed the grade placement score to indicate the level at which 50 percent of the questions could be answered. If 75 percent comprehension of the McCall-Crabbs standard was desired, 1.866 was added to the formula score. Since the Flesch Formula of 1943 had been based in part on the original Lorge Formula, it also was in error. Thus Lorge updated the Flesch Formula at this point.[40]

In 1948 Flesch revised his original formula separating it into a Reading Ease Formula and Human Interest Formula. The Reading Ease Formula was based on a count of the number of syllables in a 100-word sample and the average sentence length. The Human Interest Formula was based on a count of the personal words and personal sentences. In calculating the Reading Ease Formula the number of syllables or word length and sentence length were substituted for wl and sl respectively in the following formula:

Reading Ease = 206.835 - .846wl - 1.015sl,

where Reading Ease represented the grade level to be attained in order to read the passage. Calculation of the Human Interest factor was done by substituting personal words and personal sentences for pw and ps respectively in the following formula:

Human Interest = 3.635pw + .314ps.

Grade level equivalents were deduced from scales instead of the equations for both the Reading Ease and the Human Interest. Correlations with the

---

[40]Irving Lorge, "The Lorge and Flesch Readability Formulae: A Correction," School and Society, 67 (February 21, 1948), 141-142.

McCall-Crabbs criteria were .70 for the former and .43 for the latter.[41,42,43,44]

Dale and Chall found fault with the Flesch Formula because it was too time-consuming, due mainly to the count of affixed morphemes which was often unreliable. Thus in 1948 they developed a formula which employed a word list that would be more suitable for children's books. The basis of their efforts was a list of 3,000 words mastered by at least 80 percent of a sample of fourth grade students. The resulting two-factor formula consisted of a count of words in a 100-word passage which were not on the list of 3,000 common words and the average sentence length. The resulting data were inserted into an equation in which the Dale score was x_1 and the average sentence length was x_2. The equation was:

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

where $x_{c_{50}}$ was the reading grade score of a pupil who answered one half of a series of test questions correctly. Using the McCall-Crabbs

---

[41]Rudolf F. Flesch, "A New Readability Yardstick," Journal of Applied Psychology, 32 (June 1948), 221-233.

[42]Rudolf F. Flesch, The Art of Readable Writing (New York: Harper and Brothers Publishers, 1949), pp. 213-216.

[43]Rudolf F. Flesch, How to Test Readability (New York: Harper and Brothers Publishers, 1951), pp. 1-10.

[44]Rudolf F. Flesch, How to Write, Speak, and Think More Effectively (New York: Harper and Brothers Publishers, 1960), pp. 305-313.

passages as the criterion, the Dale-Chall Formula for readability correlated at .70.[45,46,47]

Farr, Jenkins, and Paterson in 1951 published a procedure for simplifying the Flesch Formula. The count of syllables was changed to a count of monosyllabic words. A correlation of .93 was found between results produced by the two methods on 360 selected paragraphs. The formula which resulted from this work was:

New Reading Ease Formula = 1.599nosw - 1.015sl - 31.517,

where nosw represented the number of one syllable words and sl represented average sentence length. A table was produced to speed the calculation of readability.[48]

In 1952 Robert Gunning published a new readability formula, the Fog Index. His formula required a count of polysyllabic words and average sentence length in a 100-word passage. These two factors were totaled and multiplied by .4 to obtain the Fog Index. The Fog Index yielded the reading grade level required for understanding the material.[49]

---

[45]Edgar Dale and Jeanne L. Chall, "A Formula for Predicting Readability," Educational Research Bulletin, 27 (January 1948), 11-20, 28.

[46]Edgar Dale and Jeanne L. Chall, "A Formula for Predicting Readability Instructions," Educational Research Bulletin, 27 (February 1948), 37-54.

[47]Edgar Dale and Jeanne Chall, "The Concept of Readability," Elementary English, 26 (January 1949), 19-26.

[48]James Farr, James Jenkins, and Donald Paterson, "Simplifications of Flesch Reading Ease Formula," Journal of Applied Psychology, 35 (October 1951), 333-337.

[49]Robert Gunning, The Technique of Clear Writing (New York: McGraw-Hill Book Company, 1952).

Spache in 1953 presented a readability formula primarily intended for use with primary level reading materials. The formula used indexes of average sentence length and percentage of hard words as determined by the Dale List of 769 Words. Estimated grade levels were calculated by substituting average sentence length for sl and the percentage of hard words for hw in the following formula:

Grade Level = .141sl + .086hw + .839.[50]

In 1953 Taylor offered the "cloze" procedure for measuring the reading difficulty of material. The cloze tests were constructed by deleting every nth word in a passage and replacing words with blanks of uniform length. Students filled the blanks with words they believed the author had written. The average percentage of agreement between the subjects' responses and the author's original was called the "mean cloze score" for the passage. Only those words that were replaced exactly from the original test were scored as correct responses.[51]

Bormuth in 1967 helped to clarify the cloze procedure when his research indicated:

> A student must answer correctly at least seventy-five percent of the items over a passage before the passage is said to be suitable for his use. When his score falls between seventy-five and ninety percent, the material is said to be suitable for use in supervised instruction. A score above this range is taken as evidence that the

---

[50]George Spache, "A New Readability Formula for Primary Grade Reading Materials," Elementary School Journal, 53 (March 1953), 410-413.

[51]Wilson Taylor, "Cloze Procedure: A New Tool for Measuring Readability," Journalism Quarterly, 30 (Fall, 1953), 415-433.

material may be used for independent study; a score below this range indicated that the material is too difficult for ordinary instructional purposes.[52]

In 1968 Edward Fry published a revision of his Readability Graph, which had first been published in Uganda in 1963. His formula required the calculation of the average sentence length and the average number of syllables in a 100-word passage. Approximate grade levels were found by plotting both averages on the Fry Readability Graph.[53,54]

In 1969 McLaughlin published an even simpler formula which he called SMOG. This formula depended on a count of every word of three or more syllables in thirty selected sentences. The square root was then estimated for the number of polysyllabic words in the passage. To the approximate square root he added a constant of three. The result was the SMOG grade or the reading grade level that a person must have attained in order to understand fully the material. McLaughlin found a correlation of .71 between the SMOG Formula and McCall-Crabbs Standard Test Lessons in Reading.[55]

In 1974 Spache revised his Spache Readability Formula. The new formula also included a revised word list for assessing written materials

---

[52]John Bormuth, "Comparable Cloze and Multiple Choice Comprehension Test Scores," Journal of Reading, 10 (February 1967), 292.

[53]Edward Fry, "A Readability Formula That Saves Time: Readability Graph," Journal of Reading, 11 (April 1968), 513-516, 575-578.

[54]Edward Fry, "A Readability Graph for Librarians, Part I," School Libraries, 19 (Fall, 1969), 13-16.

[55]G. H. McLaughlin, "SMOG Grading: A New Readability Formula," Journal of Reading, 11 (May 1969), 639-46.

at the primary level.[56] Burmeister in 1976 constructed a chart which would allow the investigator to by-pass the Spache Formula. In order to use the chart, the average sentence length had to be calculated and the percentage of hard words had to be determined. Using these two factors the chart would provide the approximate grade level score.[57]

Raygor devised the Raygor Readability Estimate in 1977 which was based on a graph similar to the graph developed by Fry. The major difference between the two approaches is that Raygor proposed counting long words rather than syllables. A long word is a word which has six or more letters.[58]

In 1980 Layton prepared a chart which is considered an extension of Dale and Chall's formulas for use at grade four and above. The chart was intended to encourage readability analysis by people who felt uncomfortable with the level of mathematics required by the original formula of Dale and Chall. Average sentence length and the percent of difficult words must be computed prior to entering the chart.

---

[56]George D. Spache, "The Spache Readability Formula," Good Reading for Poor Readers (Champaign, Illinois: Garrard Publishing Company, 1974), pp. 195-207.

[57]Lou E. Burmeister, "A Chart for the New Spache Formula," The Reading Teacher, 29 (January 1976), pp. 384-385.

[58]Alton L. Raygor, "The Raygor Readability Estimate: A Quick and Easy Way to Determine Difficulty," Reading: Theory, Research and Practice, The Twenty-Sixth Yearbook of the National Reading Conference, 1977, pp. 259-263.

Layton suggested that the chart was intended as a companion to the work of Burmeister which was based upon the Spache Readability Formula.[59]

Readability and Elementary School Science Textbooks

The Tennessee Department of Education has shown a great deal of concern over the reading ability of elementary students in recent years. This fact was of interest when the review of research revealed that there has been a concern over the preparation of readable materials in science with references dating to the beginning of the century.

In 1924 Curtis analyzed 630 articles containing science content which had been published in six American newspapers over a two week period. His interest centered on the different scientific terms which did not appear on Throndike's List of Ten Thousand Words. The investigation revealed that 35.1 percent of the scientific terms were found on the List of Ten Thousand Words. He concluded:

> . . . that the range of scientific vocabulary found in the public press emphasized the need for extensive training in science as a foundation for an intelligent reading of the newspapers.[60]

Curtis later studied scientific articles with particular reference to the percentage of common words, percentage of scientific terms, and

---

[59]James R. Layton, "A Chart for Computing the Dale-Chall Readability Formula above Fourth Grade Level," Journal of Reading, 23 (December 1980), pp. 239-244.

[60]Francis D. Curtis, Some Values Derived from Extensive Reading of General Science (New York: Teachers College Press, Columbia University, 1924), pp. 10-27.

percentage of uncommon words which were scientific. Uncommon words were those words which did not appear on the List of Ten Thousand Words. Among the findings were: (1) only 3.6 percent of the words used were uncommon; (2) percentage of uncommon words which appeared in the articles averaged approximately the same for all articles; and (3) scientific terms which were uncommon ranged from 23.3 percent to 38.5 percent per article.[61]

Kearney investigated the premise that sentence length was a factor in making reading material difficult to read. His attention centered on "42 pre-primers, 38 primers, and 41 first readers published during the ten years between 1930 and 1939, inclusive." The readers represented a number of different subject matter areas of which one was science. "Seven of the pre-primers contained less than 100 sentences, 22 contained from 100 to 299 sentences and thirteen contained from 300 to 499 sentences." The results obtained from studying the first readers indicated "all had over 500 sentences with 15 between 500 and 599, 23 between 1,000 and 1,499, and with 3 over 1,500." A computation of the median number of sentences per reader-type showed pre-primers with 157, primers with 775, and first readers with 1,093. Average sentence lengths for pre-primers was 0 to 6.3 words with a type average of 4.5 words. Primers had a range of 4.7 to 7.4 words with a type of average of 5.7 words, while the first readers had a range of 5.5 to 9.3 words with a type

---

[61]Francis D. Curtis, "A Study of the Vocabulary of Scientific Articles Appearing in Daily Newspapers," School and Society, 23 (June 26, 1926), 821-824.

average of 7.3 words; Kearney pointed out that children are called on to read longer sentences as they progress from pre-primers to first readers. Teachers who have studied the place of sentence length and its association with gradation of materials can more accurately select reading materials to meet the needs of students.[62]

In 1950 Mallinson, Sturm, and Patton analyzed the elementary science textbooks for grades four, five, and six for five series using the Flesch Formula. A review of the literature revealed that no studies of fourth, fifth, and sixth grade science textbooks had taken place. They concluded: (1) three of the five books written for grade four were deemed too difficult; (2) grade five textbooks were calculated to be too difficult for students at the beginning of the school year, with one textbook estimated at the sixth grade level; (3) grade six textbooks were judged only slightly difficult; and (4) there was no indication that the first sections of the textbooks were rated lower than the later sections of the textbooks.[63]

Also in 1950, Mallinson, Sturm, and Mallinson studied the readability of junior high school textbooks. This was perhaps the first readability examination of textbooks intended for grades seven, eight, and nine.

---

[62]Nolan C. Kearney, "Sentence Length in 121 Representative First-Grade Readers," Journal of Educational Research, 38 (February 1945), 447-461.

[63]George Greisen Mallinson, Harold E. Sturm, and Robert E. Patton, "The Reading Difficulty of Textbooks in Elementary Science," Elementary School Journal, 50 (April 1950), 460-463.

Only those textbook series were used in the study which were a three-book series intended for grades seven, eight, and nine and for those series in which the textbooks for grades seven and eight were part of an eight-book series. A total of twelve series was selected of which ten series were of the three-book type. The Flesch Formula was used for readability analysis. The results of the investigation were: (1) two of the textbooks intended for grade seven were too difficult for most students; (2) average level of difficulty for textbooks intended for grade eight was below that for grade seven; (3) a greater range of difficulty was found for grade nine books than for those intended for grade seven or grade eight, while only four books were rated moderately difficult for all but the better students; (4) grade seven textbooks were rated the most difficult; and (5) samples taken at the beginning of the textbooks did not seem to be less difficult than samples taken near the end of the book.[64]

Burkey in 1954 employed the Yoakam Formula to analyze textbooks intended for elementary science. A comparison of the reading levels supplied by publishers with those produced by the Yoakam technique resulted in 18 of the 41 investigated textbooks being estimated in the same grade level, while 16 textbooks were found to be above grade level and seven were below grade level. This analysis which concentrated on the use of technical terms indicated that approximately 35 percent of

---

[64]George Greisen Mallinson, Harold E. Sturm, and Lois M. Mallinson, The Reading Difficulty of Textbooks in Junior High School Science," The School Review, 58 (December 1950), 536-540.

all scaled words were technical words. A comparison between the reading difficulty of a given page and its position in the book produced an average correlation of .03, indicating there was inconsistency in gradation. Textbooks intended for grade four were found to have the most difficult reading material appearing in the front of the textbook as often as it was found in the later pages of the textbook. Burkey reported that:

> A general conclusion is that elementary science textbooks are well written, but through further research and continued improvements in vocabulary control, they will more closely fit the characteristics of the readers.[65]

Mallinson, Sturm, and Mallinson studied the reading difficulty of unit-type elementary school science textbooks. The unit-type materials studied were often pamphlet-like and centered on specific topics of science rather than the more general approach of a basal textbook. Publishers were asked by the authors to submit materials for the study and to restrict the submissions to materials useful in grades four, five, and six. The Flesch Formula was employed to determine the readability of the textbooks. The authors studied 34 publications submitted by five publishers and concluded: (1) the unit-type textbooks involved in the study were likely to cause difficulty if used at the grade level recommended by the publisher; (2) materials intended for grade four tended to have a readability higher than grade four; (3) materials intended for grades five and six were written at a level of difficulty

---

[65]Jacob E. Burkey, "The Readability of Elementary Science Materials," Dissertation Abstracts, 14 (March 1954), 1328.

closer to the publisher's recommendation; and (4) average reading levels were meaningless due to the variations observed for the individual passages studied.[66] The authors further concluded:

> It may be stated that if teachers have found textbooks of elementary science generally too difficult, they are not likely to find the unit-type textbooks much better if they use them at the grade levels that the publishers suggest.[67]

Mallinson, Sturm, and Mallinson centered their attention on elementary science textbooks again in 1957. They pointed out that earlier studies indicated:

> 1. The levels of reading difficulty of many textbooks in science are too advanced for students for whom they are written.
> 2. The differences between the levels of reading difficulty of the easiest and most difficult textbooks in all areas of science are significant.
> 3. In some textbooks of science whose average level of reading difficulty seems satisfactory, there are passages that would be difficult even for some college students.
> 4. Many textbooks of science contain non-technical words that could be replaced with easier synonyms.[68]

The authors decided to analyze more recently published textbooks with the same techniques that had been used previously to determine if level of reading difficulty had been reduced in the new textbooks. Those textbooks of all grade levels which were published during 1953 and

---

[66]George G. Mallinson, Harold E. Sturm, and Lois M. Mallinson, "The Reading Difficulty of Unit-Type Textbooks for Elementary Science," Science Education, 39 (December 1955), 410.

[67]Ibid.

[68]George Greisen Mallinson, Harold E. Sturm, and Lois M. Mallinson, "The Reading Difficulty of Some Recent Textbooks for Science," School Science and Mathematics, 57 (May 1957), 364.

1954 and had not been previously analyzed were scaled as to readability using the Flesch Formula. Mallinson et al., concluded: (1) variation abounded among the textbooks, thus indicating that reading difficulty should be a factor in textbook selection; (2) many textbooks were rated at grade levels higher than the students for whom they were intended, rather than a grade level lower as recommended; (3) average grade levels may appear to make a textbook sutiable while concealing many very difficult sections; (4) newer textbooks are as variable and as difficult as the older textbooks; and (5) authors should be warned to pay more attention to readability, since no textbook is of "much value to students if they cannot read it with sufficient ease to understand it."[69]

Denslow studied eight grade one science textbooks on the assumption that "vocabulary difficulty is probably most pronounced in first grade science texts."

> In this study, the difficulty of science textbooks is evidenced by the number of new words presented, the number of new words found only in special texts and the per cent of the total vocabulary represented by these words, comparison of different words presented with well-known vocabulary lists, and the application of the Spache Readability Formula.[70]

Among the findings were: (1) different words per textbook ranged from 52 to 253; (2) words not common with any other textbook ranged from 8 to 48; (3) average sentence length ranged from 4.3 words to 6.7 words

---

[69]Ibid., p. 366.

[70]Orriene D. Denslow, "Vocabulary and Sentence Study of Eight First Grade Science Books," Elementary English, 38 (November 1961), 487.

as compared to 5.5 words to 9.3 words as reported by Kearney; (4) textbooks with the smallest number of different words used four times or more did not always have the shortest average sentence length; (5) range of difficulty per textbook was two months to seven months; (6) average grade placement was from 1.6 to 1.9; and (7) grade placements made by authors or publishers did not "necessarily present an accurate representation of the vocabulary burden and sentence encumbrance of the text."[71]

Moore studied unit-type textbooks selected for low ability junior high school students. A survey of publishing companies, curriculum laboratories, city school systems, and selected state departments of education indicated these materials were virtually nonexistent. This finding and the demonstrated need for such materials led to an investigation of the feasibility of preparing "special science instructional materials."[72]

> The science vocabulary lists compiled by Francis D. Curtis, Luella Cole Pressey, and S. R. Powers were used in conjunction with Thorndike's Word List as resources to aid the author in writing the materials at an appropriate reading level. In order to facilitate comprehension, unfamiliar science words were defined upon introduction, redefined and repeated in several different context. . . . A concerted effort was made by the author to avoid the use of many unfamiliar non-scientific words.[73]

---

[71]Ibid., pp. 488-490.

[72]Arnold J. Moore, "The Preparation and Evaluation of Unit Text Materials in Science for Low Ability Junior High School Students," Science Education, 46 (December 1962), 488.

[73]Ibid., pp. 488-489.

"The Dale-Chall Readability Formula was used to calculate the readability of the prepared materials and three representative general science textbooks." The experimental materials and one of the three textbooks were rated grade five to grade six, while the other two textbooks were rated grade seven to grade eight. Moore concluded: (1) materials "prepared by the investigator were more effective than standard science textbooks in enhancing the low ability ninth grade student;" (2) the levels of reading difficulty of the prepared materials were more appropriate than the selected textbooks for low ability students; (3) by adjusting "vocabulary load, sentence structure, number and abstractness of concepts presented . . ." the prepared materials were easier to read than the textbooks; and (4) prepared materials were rated by science teachers as "equal to or superior to corresponding selections from standard general science textbooks with respect to scope of science concepts . . ."[74]

Rammel selected five general science textbooks which were adopted for use by the State of Indiana for study. A reading ease score was found for each textbook. The Flesch "How Easy" chart was used to convert the reading ease score to a grade level. A human interest score was also calculated for each of the five textbooks.[75] Rammel concluded:

> 1. In agreement with Mallinson, Sturm, and Mallinson, there is a considerable variation in readability, and

---

[74]Ibid., pp. 489-491.

[75]Joseph A. Rammel, "An Objective Study for the Bases of Selecting an Eighth Grade General Science Textbook," Science Education, 47 (April 1963), 258-264.

and in human interest among textbooks designed for the same grade level.

2. The level of reading difficulty for texts A, B, C, D, and E range from grade seven to grade nine.

3. Human interest among the five books ranges from "mildly interesting" to "interesting."

4. Texts A, B, C, and E are fairly easy to read and mildly interesting.

5. The highest ranking text in human interest is D (interesting).

6. If the reading difficulty of a textbook should be one grade below that for which it was designed, we should conclude that textbook D is a little difficult for the eighth grade level.

7. Authors and publishers of general science textbooks are apparently giving more attention to the interest and readability for the grade concerned.[76]

Blue investigated the influence of readability and other characteristics of science materials on the reading comprehension of seventh grade pupils. The findings were:

1. No significant difference was found in science reading comprehension between seventh grade students who read selections containing variations in readability difficulty of at least three years . . . .

2. Science reading comprehension of selection variations for seventh grade students of high and low intelligence was significantly different.

3. Data from the rating scale questionnaire suggest that seventh grade students are aware of the reading difficulty factors, an increase in sentence length and vocabulary difficulty, but these factors have little influence on pupils' comprehension of more difficult graded science material.[77]

In 1965 Ottley calculated the readability of science textbooks intended for grades four, five, and six. Four textbook series were

---

[76]Ibid., pp. 263-264.

[77]Larry Lamar Blue, "A Study of the Influence of Certain Factors in Science Materials on the Reading Comprehension of Seventh Grade Pupils," Dissertation Abstracts, 25 (April 1965), 5625.

used in the study and readability was calculated by the Lorge Formula.[78] Ottley concluded:

> 1. . . . the fourth grade science textbooks were far too difficult, for fourth grade students; the fifth grade books were slightly difficult, while the sixth grade books were the best suited for the grade for which they were intended.
> 2. Most of the books evaluated showed no reading growth. Particularly in the fifth and sixth grade, the first part of the book was not easier than the last part . . . .
> 3. Other studies on the readability of science texts for the upper elementary grades have taken place. These studies, though different formulas were used and different books were evaluated, came to the same conclusions—namely, there is a definite need for readable textbooks, particularly in grade four.[79]

Newport investigated the readability of nine textbook series written for grades one through six. The Spache Formula was employed to calculate the readability of grades one through three, while the Yoakum Formula was used for grades four through six. Newport's data indicated that the nine series were at grade level or below for all textbooks except one after consideration of textbooks intended for grades one through three. The grade four textbooks were written at grade level or below in four series and above grade level in five series. Five grade five textbooks and five grade six textbooks were also written above grade level according to the readability formula used. The largest variation was found in the grade six textbooks where three textbooks were also written two to five grades above grade level. Where

---

[78]LeRoy Ottley, "Readability of Science Textbooks for Grades Four, Five, and Six," School Science and Mathematics, 65 (April 1965), 363-366.

[79]Ibid., p. 366.

possible, Newport compared his data with data obtained by Burkey, thus allowing comparisons on four textbook series written for grades four, five, and six. A comparison of grade four textbooks showed three of the four series had lowered readability levels. The same results were found for grade five, while three series showed increases for grade six.[80]

A report of word count comparisons was made by Stauffer using seven reading series and three series each in health, science, and arithmetic. Only textbooks intended for use in grades one through three were included. His findings after study of the science series were: (1) 352 different words were used in grade one textbooks; (2) 640 different words were used in grade two textbooks; (3) 1,108 different words were used in grade three textbooks; (4) 2,100 different words were introduced in grades one through six, which was the smallest number of different words for any of the four subject areas studied; (5) significantly fewer common words appeared in the science series; and (6) science textbooks contained the fewest content words not appearing in the Master Reading Vocabulary List.[81]

Williams investigated the effects of rewriting sixth grade science textbook selections to a "third-grade level of readability on reading comprehension and reading rate of sixth-grade pupils." His data

---

[80]John F. Newport, "The Readability of Science Textbooks for Elementary Schools," Elementary School Journal, 66 (October 1965), 40-43.

[81]Russell G. Stauffer, "A Vocabulary Study Comparing Reading, Arithmetic, Health and Science Texts," The Reading Teacher, 20 (November 1966), 141-147.

indicated: (1) all three reading ability groups produced higher comprehension score means on the rewritten material; (2) low reading ability students did significantly better after studying the rewritten material than they did on the sixth grade material; (3) average and above average reading ability students performed significantly better after studying the rewritten material than they did on the sixth grade material; and (4) average and above average ability students comprehended better than low reading ability students even when the science materials were rewritten to a much lower level.

Williams recommended:

> 1. Authors and publishers of sixth-grade science textbook materials should give serious consideration to writing sixth-grade science textbooks to a much lower level of reading difficulty.
>
> 2. Science textbook selection committees should consider the factor of readability before adopting elementary science textbooks.[82,83]

Cramer and Dorsey in 1969 calculated the readability of six elementary school science series. The Spache Formula was employed for textbooks intended for grades one, two, and three, while the Dale-Chall Formula was used to determine the readability of textbooks intended for grades four, five, and six. They concluded: (1) for 22 of the 36 textbooks analyzed the estimated grade level was above the grade which

---

[82]David L. Williams, "Rewritten Science Materials and Reading Comprehension," Journal of Educational Research, 61 (January 1968), 204-206.

[83]David L. Williams, "The Effect of Rewritten Science Textbook Materials on the Reading Ability of Sixth-Grade Pupils," (Unpublished Doctoral Dissertation, University of Illinois, 1964).

had been assigned by the publisher; (2) only two of the six series had progressively more difficult reading matter as the recommended grade level progressed for textbooks intended for grades one, two, and three; (3) only four of the 18 primary books were written above grade level and all four were intended for grade one; (4) all 18 books intended for grades four, five, and six were written at grade levels different from the grade level recommended by the publisher; (5) five of the six grade four textbooks were written three to six levels above the publisher's recommendation; (6) five of the six grade five textbooks were written four to five grade levels above the publisher's recommendation; (7) all six of the grade six textbooks were written three to six grade levels above the publisher's recommendations; (8) eight of the 36 textbooks were written so that the most difficult third of the textbook was the first third; (9) 15 of the 36 textbooks were written so that they were progressively more difficult from the first third to the last third; and (10) portions of eight textbooks intended for grades four and five were calculated at the college level.[84]

Gilbert in 1973 investigated the consistency between readability levels and the appropriate grade levels for elementary science textbooks. He also examined the degree of relationship that existed between selected readability formulas when applied to identical passages taken from elementary science textbooks. The study included 10 basic science

---

[84]Ward Cramer and Suzanne Dorsey, "Science Textbooks: How Readable Are They?" Elementary School Journal, 70 (October 1969), 28-33.

series and the readability formulas presented by Spache, Dale-Chall, Lorge, and Fry. Gilbert reported:

> . . . that when the readability levels of selected science texts were considered as a group using the Spache Formula the results tended to be slightly higher than would be expected of students found in grades 1, 2, 3.
>
> . . . that when the readability levels of selected science texts were considered as a group using the Fry Formula the results tended to be consistent with expected averages and ranges of reading levels of students found in grades 1, 2, and 3.
>
> . . . that when the readability levels of selected science texts were considered as a group using the Lorge Formula the results tended to be slightly higher than the expected reading levels of students found in grades 4, 5, and 6.
>
> . . . that when the readability levels of selected science texts were considered as a group using the Fry Formula the results tended to be consistent with expected ranges of reading levels of students found in grades 4, 5, and 6. The average reading level of the passages tended to be slightly higher than should be expected in grades 4, 5, and 6.
>
> . . . that when the readability levels of selected science texts were considered as a group using the Dale-Chall Formula the results tended to be consistent with expected averages and ranges for reading levels of students found in grade 6. However, in grades 4 and 5, the average reading levels of the texts ran slightly higher and the ranges were smaller and at a slightly higher level than should be expected for these grades.[85]

In 1975 Wofford investigated the relationship between "student's ability to read and comprehend science reading material." The Reading Miscue Inventory and the Reading and Science subtests of the SRA Achievement Tests were used to collect the necessary scores from fourth and fifth grade subjects. The following significant correlations were

---

[85]Charles D. Gilbert, "An Examination of Readability Levels for Selected Basic Science Texts," School Science and Mathematics, 73 (December 1973), 747-758.

found in association with fourth grade subjects: (1) SRA Reading and Science Scores; (2) RMI Retelling and SRA Reading scores; (3) RMI Retelling and SRA Science scores; and (4) RMI Retelling and SRA Science scores. There were no significant relationships discovered for any of the variable comparisons as associated with the fifth grade subjects. Wofford concluded:

> 1. The fourth-grade text was of the correct readability level for the fourth-grade students reading on the grade level specified by the company.
> 2. The fifth-grade text was inappropriate for the fifth grade subjects because of the vocabulary and concept load.
> 3. An understanding of science content is not necessarily related to the ability to read but may be dependent upon the organization, concept and vocabulary load of the material, or the author's style.
> 4. Measures of both product and process are needed to determine the readability of material.
> 5. The Reading Miscue Inventory may be a better way of determining the readability of reading material since it measures both the reading process and the product.
> 6. Textbook companies need to search for better ways of "grade leveling" texts, especially science texts.
> 7. Schools and school divisions should consider using more than just the suggestions of textbook companies when selecting text materials.[86]

In 1975 Field et al. reviewed twelve elementary school science textbook series as to readability. Samples of 100 words were taken at twenty page intervals for each book. The Fry Readability Graph was employed to find a reading level mean and reading level range for each book in each of the twelve series. The study culminated with a tabulation of the frequency of occurrence of approximate reading levels

---

[86]Barbara A. Wofford, "Using Reading Miscue Analysis to Investigate Publishers' Suggested Readability Levels for Elementary Science Textbooks: A Comparative Study," (Unpublished Doctoral Dissertation, Virginia Polytechnic Institute and State University, 1975).

and the average approximate reading level by grade for the twelve elementary science series studied.

Consideration of the frequency of occurrence of approximate reading levels produced the following findings: (1) two grade one textbooks were written one grade below grade level, nine were written at grade level, and one was written one grade above grade level; (2) five grade two textbooks were written at one grade below grade level, and seven were written at grade level; (3) one grade three textbook was written one grade below grade level, ten were written at grade level, and one was written one grade above grade level; (4) five grade four textbooks were written at grade level, and seven were written one grade above grade level; (5) two grade five textbooks were written at grade level, eight were written one grade above grade level, and two were written at grade level; and (6) seven grade six textbooks were written at grade level, four were written one grade above grade level, and one was written two grades above grade level.

The average approximate reading level by grade for the twelve elementary science series was found to be 0.9 for grade one, 1.6 for grade two, 3.0 for grade three, 4.6 for grade four, 6.0 for grade five, and 6.5 for grade six.

Field et al. noted:

> . . . that the approximate reading level averages for the textbooks intended for grade one through grade three were less than or equal to the grade level for which they were recommended. For grade levels four through six, the approximate reading level averages were higher than the grade level for which they were intended.[87]

---

[87]Field, pp. 1-34.

In 1979 Shymansky and Yore studied six elementary school science series. They used the "Fry Readability approach and a 10 percent random sample of all reading materials from each grade level within a program . . . ." The findings of this study were:

> . . . The average reading level was observed to progress generally throughout the graded texts within each series, but each series was marked by gaps or regressions in the reading levels . . . . Dramatic jumps such as these could produce a myriad of problems for both the teacher and the students and explain some of those "unexplainable" changes in student performance and attitudes so often observed between one grade and the next.
>
> The second interesting feature revealed by the analysis was that the commonly reported averages masks extreme variation in reading level within texts supposedly specified for a given grade. For example, in Program A, grade 6 was rated at an average reading level of 6 but contained segments ranging in level from 2 to 12. This means that, in given sections in this particular text, sixth grade students might be expected to read material at a grade level as low as grade 2 or as high as grade 12. . . . the potential strain placed on poor readers using this program could be immense and lead to many complications, including a disenchantment with science.[88]

In 1979 Bennett studied the readability levels of twelve sixth grade textbooks adopted by the Utah State Textbook Commission. The Dale-Chall Formula was used to determine the readability levels. His findings showed: ". . . that two texts had a grade-level of 5-6, five texts had a grade-level of 7-8, and five texts had a grade-level of 9-10."[89]

---

[88]James A. Skymansky and Larry D. Yore, "Assessing and Using Readability of Elementary Science Texts," School Science and Mathematics, 79 (December 1979), 672.

[89]Lyle Gene Bennett, "An Analysis of the Readability of Sixth Grade Science Textbooks Using the Dale-Chall Formula and the Cloze Procedure Test," Dissertation Abstracts, 40 (April 1980), 5299a-5300A.

Bennett made the following recommendations:

> . . . that readability levels of the science texts be compared by school districts with their students reading abilities to select textbooks which have appropriate reading levels; that textbook publishers develop sixth grade science materials which are readable by a high percentage of sixth grade students; and that publishers report readability findings of their published reading materials.[90]

Four of the elementary school science textbook series studied by Shymansky and Yore had also been included in the study of Field et al. A comparison of the average readabilities produced by the two studies indicated identical results on 18 of the 24 textbooks included in the studies. The six textbooks whose readability averages were not agreed upon by the two studies differed by no more than one grade level when the respective results were compared.

Readability and Secondary School Science Textbooks

Although readability studies related to secondary school science are not important to the focus of this study they are of significance in defending the use of readability in the discipline. Studies of readability with science textbooks prior to 1950 were for the most part associated with secondary school science. The presentation which follows includes only selected studies considered to be important for a chronological history of readability formulas used at the secondary school level.

---

[90] Ibid., p. 5300A.

The most often referenced material was prepared by Curtis in 1938.[91] Curtis wrote in 1969 that, ". . . the problem of reading difficulty, with its attendant and related problems, has long been recognized as one of the most important in the entire field of education."[92] His studies centered on word counts which were dependent upon the works of Thorndike. Curtis pointed out, ". . .there seems to be a need, therefore, for vocabulary investigations of textbooks of science extensive enough to be significant, which by employing some word counts will avoid objections raised to conclusions based on sampling techniques."[93] His data bank consisted of the results of 100 studies. These included 55 master's theses which had been prepared under his supervision at The University of Michigan. The studies dealt with four major areas of vocabulary investigation.

A series of 18 studies was conducted between 1930 and 1931. "The purpose of this investigation was to establish a practicable, valid, and reliable technique by which to determine pupil comprehension of the meanings of words encountered when reading prose."[94] The data was collected by having students underline words with which they were not familiar. After the investigations Curtis concluded:

---

[91]Francis D. Curtis, Investigations of Vocabulary in Textbooks of Science for Secondary Schools (Boston: Ginn and Co., 1938).

[92]Francis D. Curtis, "Investigations of Vocabulary in Textbooks of Science for Secondary Schools," Research in Science Education 1938 through 1947, Robert W. Boenig, Editor (New York: Teachers College Press, Columbia University, 1979), p. 65.

[93]Ibid.

[94]Ibid., p. 66.

1. "The method of determining vocabulary comprehension by having pupils underline the words in context which they think they do not comprehend is a convenient means of determining words which the pupils do not comprehend." This seems to be the only practical method for use with extensive selections of reading material.
2. "The underlining test possesses a high degree of reliability and validity."
3. Although there is some question about whether the underlining test can be used to determine all the uncomprehended words in a given selection, the evidence "seems conclusive . . . that such words as a pupil does underline in context as unknown to him are likely to be outside his vocabulary range."[95]

A second series of investigations involved the work of 44 graduate students conducted between 1932 and 1937. The purpose of the investigation was to determine the relationship between the technical and non-technical words in science textbooks not comprehended by junior and senior high school students and words included on Thorndike's Test of Twenty Thousand Words. Again the underlining technique was used and the collected data was compared to the 1000 word levels comprising Thorndike's List. Curtis concluded:

1. Pupils of both junior-high-school and senior-high-school levels find in their textbooks of science many technical and non-technical words which are outside their familiar vocabularies.
2. Generally speaking, many more of these unfamiliar and unknown words "belong in the seven thousand-word level of Thorndike's List than belong in any preceding level." It is felt therefore, that this should be the "logical starting point" for simplification of vocabularies by authors of science textbooks for secondary school use.
3. The first ten thousand-word levels of Thorndike's List contain a "large percentage" of the words which the pupils feel they do not know. There is an especially "high concentration" of these words in the "seventh, eighth, and

---

[95]Ibid.

ninth thousand-word levels." This seems to indicate that the authors of textbooks overestimate the vocabularies of their readers.

4. The fact that most of the words are within rather than above the list indicates at least some effort on the part of the authors "to avoid using very uncommon words.

5. Junior-high-school pupils have more difficulties with non-scientific words. Senior-high-school pupils have almost as many difficulties with the scientific and technical words as they do with non-scientific words. This could indicate to authors of books "a logical and practicable means for reducing substantially the vocabulary difficulty in" their books. This would be done simply by reducing the number of difficult non-technical terms in writings.[96]

Efforts in 1936 and 1937 produced another 30 studies. "The purpose of this investigation was to determine whether the vocabularies used in textbooks of high-school science are suitable with respect to difficulty to the pupils for whom the books were written."[97] Suitability was defined by the previous two investigations. There were 29 textbooks studied which involved six general science, six biology, eight chemistry, and nine physics. The frequency of occurrence of difficult words was calculated and percentages were computed to indicate the extent of difficult words which were scientific terms. Curtis concluded.

1. The vocabularies of these books are too difficult. They contain many words "which are rarely encountered even in literature intended for general adult reading."

2. "The vocabulary levels at which pupils . . . encounter difficulty are lower than those determined or assumed by earlier investigations." They are also lower than the textbook authors seem to anticipate.

3. "The provision for repetition is inadequate." This may be stated even though no actual determination of the

---

[96]Ibid., p. 68.

[97]Ibid., p. 69.

proper number of repetitions to insure mastery has been made. When a pupil encounters a difficult word only "once or twice" in the entire text, it is certainly not enough to properly familiarize him with that word. No repetition at all is difficult to justify.

4. "The difficulty of the vocabulary of textbooks of science is increased to an unjustifiable extent by the use of many difficult words which are not scientific terms."

5. Too few of the scientific terms found in the books are actually defined. The problem is further complicated by the fact that "scientific terms are not infrequently introduced into the textbook of science at points considerably in advance of those in which their definitions appear."[98]

Six investigators contributed to the fourth investigation. The purpose of this investigation was to determine the scientific terms which should be included in the glossaries of textbooks of general science, of biology, and of chemistry. The investigators examined six textbooks from the field of general science and six textbooks from the field of biology in one phase of the investigation. Word lists were prepared and judged as to their suitability for inclusion in a glossary. Curtis summarized:

1. Twenty-five and two-tenths percent of the general science terms and 27.4 percent of the biological terms were evaluated as essential by all six judges who evaluated them.

2. There was widespread disagreement among the authors of general science texts and between the authors and the evaluator about which terms should be considered essential.

3. Agreement was closer among authors and evaluators of textbooks of biology.

4. Basic list for inclusion in glossaries of general science books were compiled. It contained 629 words.

5. A comparable word list for biology contained 825 words.[99]

---

[98]Ibid., pp. 69-70.

[99]Ibid., pp. 70-71.

The second phase involved the investigation of 31 chemistry textbooks written between 1925 and 1935. None of the textbooks studied had a glossary; however, "three of them had lists of important words or terms, and there were definitions in some of the texts."[100] A glossary of 543 terms was prepared using the information available. This list was then presented to a panel of 40 judges for evaluation.

Between 1950 and 1954 Mallinson, Sturm, and Mallinson dominated the investigations involving the reading difficulty of science textbooks. They selected for study textbooks used to teach junior high science,[101] general science,[102] biology,[103] chemistry,[104] physics,[105] and general science and earth science.[106] In each study the researchers applied

---

[100]Ibid., p. 71.

[101]George G. Mallinson, Harold E. Sturm, and Lois M. Mallinson, "The Reading Difficulty of Textbooks in Junior High School Science," School Review, 50 (December 1950), 536-540.

[102]George G. Mallinson, Harold E. Sturm, and Lois M. Mallinson, "The Reading Difficulty of Textbooks for General Science," School Review, 40 (February 1952), 94-98.

[103]George G. Mallinson, Harold E. Sturm, and Lois M. Mallinson, "The Reading Difficulty of Textbooks in High School Biology," American Biology Teacher, 12 (November 1950), 151-156.

[104]George C. Mallinson, Harold E. Sturm, and Lois M. Mallinson, "The Reading Difficulty of Textbooks for High School Chemistry," Journal of Chemical Education, 29 (December 1952), 629-631.

[105]George G. Mallinson, Harold E. Sturm, and Lois M. Mallinson, "The Reading Difficulty of Textbooks for High School Physics," Science Education, 36 (February 1952), 19-23.

[106]George G. Mallinson, Harold E. Sturm, and Lois M. Mallinson, "The Reading Difficulty of Textbooks for General Physical Science and Earth Science," School Science and Mathematics, 54 (November 1954), 612-616.

the Flesch Formula to passages selected from the science textbooks at the rate of one passage per 100 pages with a minimum of five samples per textbook. Mallinson, Sturm, and Mallinson concluded:

> 1. The reading levels of many textbooks in science are too advanced for students for whom they are written.
> 2. The differences between the levels of reading difficulty of the easiest and most difficult textbooks in any area of science are significant.
> 3. In some textbooks of science whose average level of reading difficulty seems satisfactory, there are passages that would be difficult even for some college students.
> 4. Many textbooks of science contain non-technical words that could be replaced with easier synonyms.[107]

In 1961 Belden and Lee studied five biology textbooks as to reading difficulty using the Dale-Chall Formula and determined the level of achievement in reading of high school biology students. The conclusions reached were: (1) there is a need for greater attention to readability in the selection of biology textbooks; (2) there is a need to know the reading ability of the students so that the textbook and the learner can more nearly be matched; (3) only one of the five textbooks was considered useful to over half the students; and (4) poor vocabulary scores of the students suggested problems with the technical vocabulary of science textbooks.[108] Belden and Lee further pointed out:

> It should be emphasized that the main implication of this study lies not with the specific results but with the process involved. Other textbooks and other groups of students might

---

[107]George G. Mallinson et al., "The Reading Difficulty of Science Textbooks," Research in Science Education 1948 through 1952, J. Nathan Swift, Editor (New York: Teachers College Press, Columbia University, 1969), p. 91.

[108]Bernard R. Belden and Wayne D. Lee, "Readability of Biology Textbooks and the Reading Ability of Biology Students," School Science and Mathematics, 61 (December 1961), 689-693.

> give different results, but the need for evaluation of both the readability and the reading ability of the students for whom the books are intended obviously exists.[109]

Belden and Lee in 1962 focused their attention on five chemistry and five physics textbooks. They again used the Dale-Chall Formula and employed the formula at 50-page intervals. In addition, students from six high schools were given the Nelson-Denny Reading Test.[110] The conclusions drawn by the investigators were:

> . . . none of the chemistry textbooks had a readability score that would make it useful to even half of the students in the survey.
>
> The results relative to the physics books indicated that several of the textbooks could be used to good advantage by most of the students . . . .
>
> The main implication of this study lies in the process involved, not in the specific results . . . . One point is clear: the necessity for including the readability of textbooks and the reading ability of the students as part of the criteria for selecting textbooks.[111]

Lee and Hislop studied the readability of biology textbooks prepared for the Biological Science Curriculum Study. They compared the reading level of BSCS biology textbooks to Belden and Lee's data in readability. Specifically they used the general biology textbooks with the highest and lowest readabilities for comparison. The analysis was made using the Dale-Chall Readability Formula at fifty page intervals

---

[109]Ibid., p. 693.

[110]Bernard R. Belden and Wayne D. Lee, "Textbook Readability and Reading Ability of Science Students," The Science Teacher, 29 (April 1962), 20-21, 23.

[111]Ibid., p. 21.

on the three BSCS first year textbooks and one second year textbook.[112] The investigators concluded:

> If the compilers of the BSCS text materials designed their work for the average and above average student, the results of this study indicate that the average student may experience a great deal of difficulty in dealing with the vocabulary and subsequent concepts to be learned in the BSCS textbooks as well as from other general biology texts.[113]

Readability studies conducted on secondary school science textbooks in recent years have been fewer in number. The majority of these studies have involved small numbers of textbooks or have concentrated on a single textbook.

Comparison of Readability Formulas

Research involving comparative studies on the available readability formulas is not plentiful. At best most studies which were located dealt with a small number of different formulas.

In one of the earlier studies, Herrington and Mallinson determined that the Flesch, Lorge and Dale-Chall formulas were the most widely used formulas in 1958 based upon a review of science education publications. The passages chosen to compare the three formulas came from materials intended for grades four through eight. Each passage was subjected to readability formula analysis and to analysis by reading

---

[112]Wayne D. Lee and Madge E. Hislop, "Problems Involved in Placement of Students in a Biology Class," School Science and Mathematics, 68 (June 1968), 473-476.

[113]Ibid., p. 475.

experts. The result was an estimated readability level for each passage using each of the four sources.[114]

This study provided an opportunity to make comparisons among formulas while answering the charge that reading experts were more accurate than the formulas. The conclusions for this study were:

> 1. The data obtained in this study indicate that there is a great difference between the consistency with which the readability formulas evaluate the grade level of reading difficulty of the samples, and the consistency with which the reading experts evaluate the grade level of reading difficulty of the samples.
> 2. In none of the one hundred and ninety-nine samples did the reading experts agree completely with respect to the grade level of difficulty. However, in twenty of the samples the readability formulas gave the same measurement of reading difficulty.
> 3. In forty-nine of the sample passages the measurements made with the readability formulas differed by only .5 of a grade level. In none of the samples did the estimates of the reading experts differ by only .5 of a grade level.
> 4. In the measurement of seventy-five sample passages, the measurements made with the readability formulas differed by one grade level.
> 5. In the measurement of twenty-four sample passages, the readability formulas differed in grade level evaulation by 1.5 grade levels. The reading experts differed in their estimates by 1.5 grade levels on twelve sample passages.
> 6. In the evaluation of eighteen sample passages the readability formulas differed by two grade levels, as compared with twenty-five samples in which the estimates of reading experts differed by two grade levels.
> 7. In the measurement of six samples, the readability formulas varied by 2.5 grade levels. In the estimates of twenty-six samples, the reading experts differed by 2.5 grade levels.
> 8. In only two of the samples measured by the readability formulas did the evaluations differ by three grade levels.

---

[114]Roma Lenore Herrington and George Greisen Mallinson, "An Investigation of Two Methods of Measuring the Reading Difficulty of Materials for Elementary Science," Science Education, 42 (December 1958), pp. 387-388.

However, the estimates of reading difficulty varied by three grade levels in fifty-one samples estimated by reading experts.

9. The measurement of grade level of difficulty by readability formulas differed by 3.5 grade levels in three samples. The estimates of grade level of difficulty by reading experts differed by 3.5 grade levels in twenty-two samples.

10. The measurements by formulas differed by four grade levels in one sample analyzed, as compared with thirty-seven samples in which the estimates of reading experts differed by four grade levels.

11. The measurements by formulas differed by 4.5 grade levels in one sample. The estimates of reading experts differed by 4.5 grade levels in seventeen samples.

12. In none of the samples measured did the readability formulas differ by five grade levels. However, the estimates of reading experts differed by five grade levels in four samples.

13. In none of the samples analyzed did the readability formulas differ in measurement by 5.5 grade levels. In two of the samples the estimates of the reading experts varied by 5.5 grade levels.

14. The data obtained by the study indicate that readability formulas measure grade levels of reading difficulty of science materials with much more consistency than reading experts estimate grade levels of reading difficulty.

15. The data reveal no great difference in the consistency of measurement by the readability formulas at various grade levels. However, the measurements were slightly less consistent at the seventh- and eight-grade levels.

16. The estimates of the reading experts were most consistent at the fourth-grade level and least consistent at the seventh- and eighth-grade levels.

17. The evaluations made for higher level materials by both formulas and experts tend to be somewhat less consistent than those for lower level materials.

18. It is apparent that the statement made by Dr. Betts that any good teacher can tell the difficulty of reading materials by examining them is not supported by the data obtained in this study. Rather, the data reveal that even those teachers who are considered reading experts are very inconsistent in their estimates of reading difficulty.[115]

---

[115]Ibid., pp. 389-390.

In 1965 Brown did a follow-up study to Herrington and Mallinson, but limited his efforts to the Dale-Chall Formula. After including the Belden and Lee study on physics textbooks into his analysis, he was led to hypothesize that:

> When applied to science books, the Dale-Chall Formula yields results which are consistently higher than the actual reading level of the books. It is further hypothesized that his high estimate is largely due to the over-abundance of "technical vocabulary" found in science books and the absence of such words from the Dale list.[116]

In 1968 Fry presented a revision of the Readability Graph. In connection with the presentation he cited correlations obtained between the Readability Graph and selected readability formulas. The results indicated correlations of .94 with the Dale-Chall, .96 with the Flesch, and .98 with the SRA, and .78 with the Botel. All of these correlations were significant at the .01 level. The Readability Graph was also deemed in agreement with the Spache Formula at the lower grade levels although no correlations were cited.[117]

Fry offered the following conclusions:

> Readability formulas have had widespread, long-term interest among professionals in the reading business. However, their lack of use in the broader educational circles may be due to excessive working time and difficulty in computing some existing formulas. The Readability Graph is presented as a faster and simpler method of determining readability. It correlates highly with the Dale-Chall, SRA, Flesch, and Spache formulas.[118]

---

[116]Walter R. Brown, "Science Textbook Selection and the Dale-Chall Formula," School Science and Mathematics, 65 (February 1965), p. 166.

[117]Edward Fry, "A Readability Formula That Saves Time," Journal of Reading, 11 (April 1968), 576.

[118]Ibid., pp. 576-577.

Stephens in 1971 used cloze tests constructed from 250 word passages in elementary science textbooks as a standard for judging the Dale-Chall Formula, a revised Dale-Chall Formula, and the Fry Readability Graph. His findings led to the following conclusions:

> 1. While readability estimates as determined by the Dale-Chall formula and the revised Dale-Chall formula were more in agreement with the results of the cloze tests than were the readability estimates of the Fry formula, none of the three formulas appears to provide a valid estimate of the reading difficulty of passages taken from elementary science textbooks.
> 2. The significant correlations found between the cloze test results and the grade level estimates as determined by the Dale-Chall formula and the revised Dale-Chall formula indicated an overall agreement in the relative difficulty of the passages but were not meaningful in terms of the grade level placement of the passages.[119]

In 1972 Driver determined the intercorrelations among the Dale-Chall Readability Formula, the Fry Readability Graph Formula, and the SMOG Grading Readability Formula. The study centered on fourth, fifth and sixth grade science textbooks. The grade levels designated by the publishers of the elementary science textbooks were also a part of the study. Driver concluded:

> 1. Significant relationships beyond the .01 level of significance were found between the following variables: Dale-Chall readability scores and publishers' designated grade levels; Fry Graph readability scores and publishers' designated grade levels; SMOG Grading readability scores and publishers' designated grade levels.
> 2. Significant relationships beyond the .001 level of significance were found between the following variables:

---

[119]Robert Gerald Stephens, "Use of the Cloze Procedure as a Criterion for Evaluating the Applicability of Selected Readability Formulas to Science Reading Materials," Dissertation Abstracts, 32 (May 1972), 6020A.

Dale-Chall readability scores and SMOG Grading readability scores; Fry Graph readability scores and SMOG Grading readability scores.

3. The Dale-Chall Readability Formula rated all of the textbooks from one to four grades above the assigned grade levels.

4. Grade level variations within each textbook, as determined by application of the Dale-Chall Readability Formula, ranged from fourth grade level to college graduate level.

5. The Fry Readability Graph rated the science textbooks nearer the publishers' assigned grade levels than did the Dale-Chall Formula or the SMOG Grading Formula.

6. The SMOG Grading Formula rated the science textbooks as being more difficult than did the Dale-Chall Formula or the Fry Graph Formula.

7. The Fry Graph and SMOG Grading Formulas were found to be simple to calculate and great time savers when compared with the Dale-Chall Readability Formula.

8. Validity of the two recently devised formulas—the Fry Graph and SMOG Grading—was established by high correlation coefficients with the older Dale-Chall Readability Formula.

9. The most important advantage of the Fry Readability Graph Formula and the SMOG Grading Formula is that they test technical vocabulary by the polysyllable count and do not depend upon the familiarity of the word.

10. Ease and simplicity of calculations make the Fry Graph Formula and the SMOG Grading Formula appropriate for teachers who do not have the time or training to administer the older formulas.[120]

In 1974 Kennedy recognized the need for teachers to determine the reading level of a student and then the reading level of the textbook that the student would be expected to read. In particular he emphasized the difficulty of determining the readability of the textbook because

---

[120]Jean Ricks Driver, "The Intercorrelations among Three Readability Formulas When Applied to Selected Fourth, Fifth, and Sixth Grade Science Textbooks," Dissertation Abstracts, 32 (May 1972), 5999A.

"reading level formulas are not convenient for the classroom teacher to apply."[121]

Kennedy's discussion included the complexity of calculation involved with available formulas, as well as the time consuming nature of the task. He further points out that "only in the last few years have relatively simple methods for determining reading level become available. One of these, the Readability Graph, was developed by Edward Fry." His justification for use of the Fry Readability Graph was based on correlations between the Readability Graph and other readability formulas which had been cited by Fry in 1968.[122]

In 1978 Carter and Simpson repeated the same outline which Kennedy had proposed. They suggested "as a first step in spotting students who need help, teachers should examine individual results of standardized tests . . . ." They further recommended, "The Fry Readability formula is one technique commonly used to evaluate the average reading level of instructional material." As a follow-up procedure to using the Fry Readability Graph in the classroom setting the authors suggested that the cloze formula also be employed.[123]

Ricker in 1978 investigated three readability formulas. The formulas studied were the Fry Readability Graph, the Fog Index, and

---

[121]Keith Kennedy, "Reading Level Determination for Selected Texts," The Science Teacher, 41 (March 1974), 26.

[122]Ibid., pp. 26-27.

[123]Glenda S. Carter and Ronald D. Simpson, "Science and Reading: A Basic Duo," The Science Teacher, 45 (March 1978), 19-20.

the SMOG Grading Formula. In order to compare the formulas Ricker used six intermediate grade level science textbooks. His justification of the variations in readability computed for the textbooks was explained using two of the textbooks which had been designated as A and B.[124]

Textbook B, for example, had a seventh grade reading level as measured by the Fry Formula, a ninth grade level as measured by the SMOG, and an eleventh grade level as measured by Fog. Textbook reading level by grade varied from 6 to 9 depending on which formula was applied. The Fry scores were consistently lower than the other two; there was no consistent difference between SMOG and Fog scores.

The first point one must keep in mind in evaluating these scores was that they were based on different prediction criteria. That is, the Fry score of 6 for textbook A predicted that a student who read at a sixth grade reading level would be able to read the text with 50 to 75 percent comprehension (the Dale-Chall Readability Formula, another popular formula, also uses this percent comprehension factor). Both the Fog Index and the SMOG Formula on the other hand, attempted to predict the reading grade level necessary to read with 90 to 100 percent comprehension. Thus, for example, the Fog Index predicted that a youngster would need an eighth grade reading level to read text A with 90 to 100 percent comprehension, the SMOG Formula predicted that a ninth-grade reading level would be necessary.[125]

---

[124]Kenneth S. Ricker, "But Can They Read? A Look at Readability Formulas," The Science Teacher, 45 (March 1978), 23.

[125]Ibid.

Ricker concluded that the use of the textbook in the classroom setting may determine the best formula to be used. Based on the above related discussion the SMOG Grading Formula or the Fog Index would be most desirable if the students were expected to perform at an independent level. Assuming the teacher provided students with a large amount of help in understanding the textbook, the Fry Readability Graph would be a suitable formula for use.[126]

In 1979 Grove compared four readability formulas. The four formulas considered were the Flesch Reading Ease Formula, Dale-Chall Readability Formula, Fry Readability Graph, and SMOG Grading Formula. The investigation involved the standard sampling procedures recommended by the respective authors of the formulas plus three modified sampling procedures. The study made use of twenty-one selected textbooks.[127]

The investigation projected the following conclusions:

> 1. . . . the classroom teacher can obtain sufficiently accurate results by using any of the more time-efficient, modified sampling procedures.
> 2. . . . either the Fry or SMOG formula can be used with as much accuracy as the more time-consuming Flesch or Dale-Chall formula.
> 3. When the comprehension criteria applied to each formula by its author(s) are taken into consideration, the ranking of the grade equivalent scores for the formula would be expected to be highest for the Dale-Chall with the Fry, Flesch, and SMOG progressively lesser. However, this ranking was not found in this study.[128]

---

[126]Ibid., p. 24.

[127]Martha Judin Grove, "Sample Frequency Modifications and Comparative Readability of Four Readability Formulas as Applied to Selected Textbooks on the Teaching of Reading," Dissertation Abstracts, 40 (September 1979), 1377A.

[128]Ibid.

Baldwin and Kaufman in 1979 compared the Fry Readability Graph with the Raygor Readability Estimate. The authors conceded at the outset that

> . . . the Fry Readability Graph is probably the most widely used readability measure. It is procedurally simple and renders readability estimates which are quite similar to those of more cumbersome formulas such as the Dale-Chall.[129]

The contention of this study was that the Raygor Readability Estimate would be just as accurate and even less time consuming.

In the study one hundred books were selected for use which ranged from third grade level through professional level. Three 100-word passages were selected from each book for analysis with one passage each taken from the front, middle and end of the book. A second phase of the study involved the calculation of the time it took to perform each of the procedures.[130]

The authors concluded:

> 1. The Raygor graph is easier to use than the Fry graph . . . . The Raygor graph could make readability assessments more accurate and more accessible to content teachers who have neither the background nor the time to become good syllable counters . . . .
> 2. The Raygor graph is considerably faster to use than the Fry graph . . . . This is an obvious advantage for teachers who are comfortable with the graph approach to readability and who have large numbers of books and other materials to review.
> 3. The Raygor graph does not appear to sacrifice accuracy for simplicity and time. The high correlation between Raygor and Fry estimates in the present study

---

[129]R. Scott Baldwin and Rhonda K. Kaufman, "A Concurrent Validity Study of the Raygor Readability Estimate, Journal of Reading, 22 (November 1979), 148-149.

[130]Ibid., p. 149.

> suggest that, for all practical purposes, the two graphs yield the same results.[131]

In 1980 Stevens questioned the validity of a number of the readability formulas. Those formulas about which doubt was raised were formulas which had been based on the McCall-Crabbs Standard Test Lessons in Reading.[132]

> Among the formulae that use the Test Lessons as the criterion of difficulty are the Lorge formula and its revision, the Flesch formula and its revisions, the Dale-Chall formula and its revisions, the Farr-Jenkins-Paterson formula, the Fog Index, the Danielson-Bryan formula, and the SMOG grading.[133]

The above listed formulas were based on the McCall-Crabbs Standard Test Lessons on the assumption that the McCall-Craggs Standard Test Lessons produced accurate grade scores. Stevens, however, had obtained documentation that the test lessons were intended "only to be practice exercises in reality."

> The grade level scores were rough equivalents, provided for students to trace their progress. The lessons were never meant to be used for vigorous testing, and were never intended for use as a criterion for readability formulae.[134]

An important note associated with this review is that only the Spache Formula, Fry Readability Graph, and the Raygor Readability

---

[131]Ibid., pp. 152-153.

[132]Kathleen C. Stevens, "Readability Formulae and McCall-Crabbs Standard Test Lessons in Reading," Journal of Reading, 23 (January 1980), 413.

[133]Ibid.

[134]Ibid., p. 414.

Estimate number among the commonly used readability formulas which were not based on the McCall-Crabbs Standard Test Lessons.

In 1980 Harris and Jacobson compared the Fry, Spache, and Harris-Jacobson readability formulas. Harris and Jacobson concluded:

> The results of this study challenge the suitability of the Fry graph for measuring the readability of primary grade reading materials. Questions have been raised about the absence of the kinds of information concerning constructive and statistical characteristics that an educational measuring instrument should provide. The main objective, however, is based on evidence that the Fry graph seriously overestimates the difficulty of second and third grade reading materials and is way out of line as compared to two other readability procedures as well as the level designations provided by the publishers.
>
> The Spache and H-J formulas are quite similar in variables used, the way the formulas were developed, and many of their statistical characteristics. The main difference discovered is that the Spache tends to underestimate the difficulty of new second and third-grade basal readers, using publishers' designations as the criterion.[135]

---

[135]Albert J. Harris and Milton D. Jacobson, "A Comparison of the Fog, Spache, and Harris-Jacobson Readability Formulas for Primary Grades," The Reading Teacher, 32 (May 1980), 923-924.

CHAPTER III

PROCEDURES

This study was designed to investigate some of the effects of elementary school science textbook readability levels on science achievement of elementary school students with low, average, and high reading abilities as measured by the Metropolitan Achievement Tests: Science.

I. SELECTION OF ELEMENTARY SCHOOL SCIENCE TEXTBOOK SERIES

Readability Formula Selection

Chapter II included major investigations of the available readability formulas and readability formulas employed in readability studies conducted on elementary science textbooks. Study of these investigations led to the selection of the Fry Readability Graph as the readability formula to be used in this study. Major reasons for this selection involved the frequent use of this formula by other researchers, consistency of results when compared with other available formulas, and the minimum effort required to employ the formula. The latter two reasons were seen as important due to the limited time of the public school practitioner who may find further action research beneficial.

Employment of the Fry Readability Graph made necessary the calculation of the number of sentences to the nearest tenth and the

number of syllables for a minimum of three 100-word samples, with proper nouns omitted.[1] Elementary school science textbooks under consideration for this study were subjected to readability samples taken at approximately 20-page intervals. This decision was made to improve the accuracy of the readability estimations which were taken as compared to Fry's recommendation to take only three samples for analysis.

Each sample taken contained 100 words, omitting proper nouns, starting with the first complete sentence of the designated page. Samples were taken from pages 1, 20, 40, etc., excluding pictures, student exercises, and student problems. A sampling that should have started on page 40 may have actually started on page 45, with the following sampling being attempted on page 60.

After all samples had been computed for a particular textbook, averages were found for the number of sentences and the number of syllables per 100 words. As stated earlier, the number of sentences per 100 words were computed to the nearest tenth. By plotting the number of sentences per 100 words and the number of syllables per 100 words on the Fry Readability Graph, the approximate reading level of the textbook was determined. The approximate reading level was computed as a whole number.

The Fry Readability Graph is presented on the following page as Figure 1. This reproduction was possible by permission of the author.[2]

---

[1]Edward Fry, Reading Instruction for Classroom and Clinic, (New York: McGraw-Hill Book Company, 1972), pp. 231-232.

[2]Ibid., p. 232.

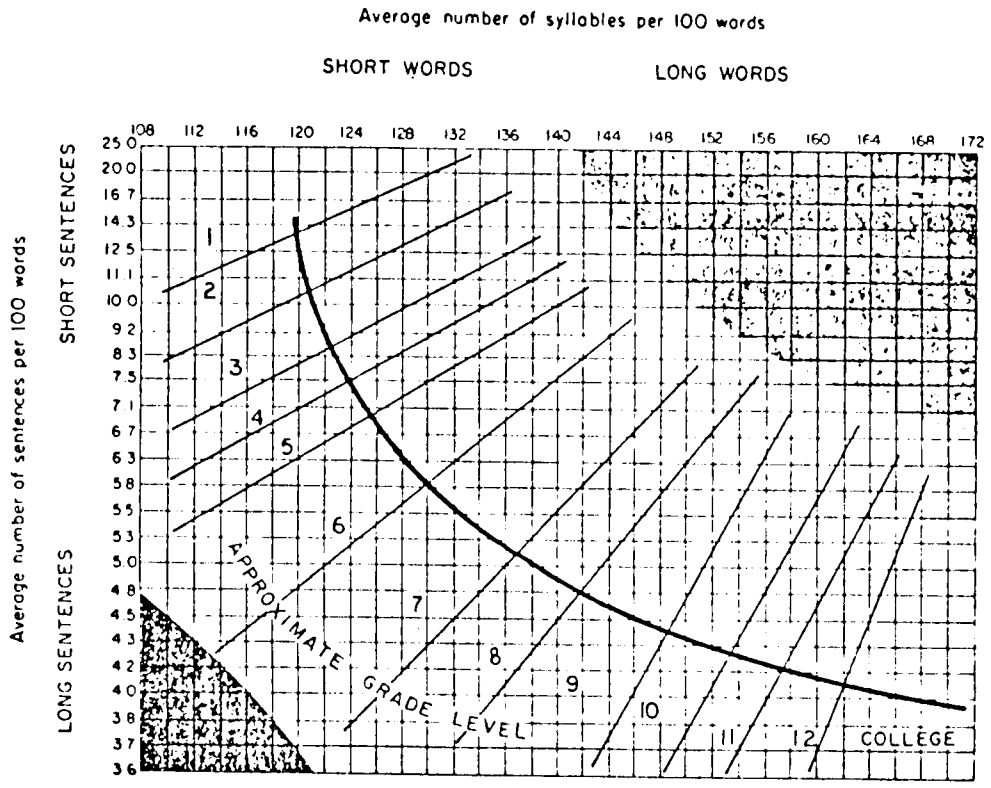


Figure 1. Fry Readability Graph

Criteria for Textbook Series Selection

A review of numerous investigations involving the readability levels of elementary school science textbook series indicated that most series have estimated readability levels close to the grade level for which the textbooks were intended for grades one through three. Textbooks intended for the upper elementary school grade levels tended to have estimated readability levels which were higher than the grade level for which the textbooks were intended.

The latter category of textbooks was the focal point for this study. The elementary school science textbook series selected for study included fifth and sixth grade elementary school science textbooks which had estimated readability levels which were higher than the grade level for which the books were intended as determined by the Fry Readability Graph.

The elementary school science textbook series selected for consideration were adopted in Tennessee in February 1975. These series were used between September 1975 and June 1980, in accordance with the five-year textbook adoption program of the Tennessee Department of Education.[3] Table I includes the elementary school science textbook series considered and the estimated readability levels of the respective fifth and sixth grade level textbooks.

---

[3]Tennessee Textbook Commission, Tennessee Official List of Textbooks with Wholesale Prices, Retail Prices and Copyright Dates, (Nashville; Tennessee Book Company, 1979), pp. 36-51.

TABLE I

Estimated Readability Levels of Fifth and Sixth Grade Elementary School Science Textbooks Adopted in Tennessee for September 1975 through June 1980

Publisher	Estimated Readability Level	
	Fifth Grade Textbook	Sixth Grade Textbook
Addison-Wesley Publishing Company	5	6
American Book Company	6	7
Ginn and Company	7	7
Harcourt Brace Jovanovich, Inc.	6	6
Harper and Row Publishers, Inc.	6	6
D. C. Heath and Company	6	6
Houghton Mifflin Company	5	7
Laidlaw Brothers	6	8
Charles E. Merrill Publishing Company	6	6
Rand McNally and Company	7	6
Silver Burdett Company	6	7
Webster Division-McGraw Hill Book Company	6	6

Source: Maruice Houston Field et al., "The Approximate Reading Level Ranges and Means by Grade Level for Twelve Elementary School Science Textbook Series," Science Education Bulletin 101 (Martin, Tennessee: Department of Elementary and Secondary Education, The University of Tennessee at Martin, 1975), pp. 1-34.

Those elementary school science textbook series including fifth and sixth grade level textbooks consistent with the above stated criteria were: American Book Company, Ginn and Company, Laidlaw Brothers, Rand McNally and Company, and Silver Burdett Company. The Silver Burdett Series was chosen for study. This textbook was adopted by the school system which met the achievement test criteria outlined in the next section. This school system later agreed to cooperate in the study.

II. SELECTION OF ACHIEVEMENT TESTS

The Metropolitan Achievement Tests were chosen for measuring both reading and science achievement. This decision was based, in part, on the frequent use of the tests in West Tennessee and the common use of these instruments for achievement test data released by the Tennessee Department of Education. These two reasons were deemed important due to the need for achievement test data collected over a three year period. Use of this annually collected data presented no additional burden to the student beyond that already considered important by the school system.

Data obtained from the Metropolitan Achievement Tests were appropriate for use in this study since both reading achievement and science achievement were measured. Specifically, reading achievement data were needed at the fourth, fifth, and sixth grade levels for the studied groups. Science achievement data were needed at the fifth and sixth grade levels. The Metropolitan Achievement Tests: Elementary Form normally administered at the fifth and sixth grade levels contained subparts which measured both reading achievement and science achievement.

III. SELECTION OF READING ABILITY GROUPS

A school system which met the previously stated criteria was selected in Northwest Tennessee. The school system adopted a basal elementary school science textbook series in 1975 which included fifth and sixth grade textbooks with estimated readability levels higher than the grade level for which they were intended. The textbook series selected was published by Silver Burdett Company.

The school system administered the Metropolitan Achievement Tests to all fourth, fifth, and sixth grade students during the school years 1977-78, 1978-79, and 1979-80. The administration of the Metropolitan Achievement Tests took place in the month of October during the two former school years and in April during the latter school year. This testing format enhanced the study since an early fourth grade reading achievement score was desirable. A late sixth grade score was also more desirable than either an early sixth grade or an early seventh grade score.

The school system selected was a metropolitan type which included all of the schools within the county. In order to limit the variables associated with teaching styles, students attending elementary school in a single town were chosen for the study. These students attended science classes in the fifth and sixth grades which were departmentalized.

Collection of the Metropolitan Achievement Tests scores used in this study was by permission of the superintendent of schools and with the aid of the elementary school supervisor. Pertinent scores were transferred from group print-outs to a specially prepared data collection

form. This transfer was made in the superintendent of school's office and did not involve in any way individual files of students.

The data collection form is exhibited in Appendix A. This form included a place to record a student's name to insure accurate recording of scores from three print-outs and to make possible verification of the data collection process, if necessary. After all scores had been recorded, the school psychologist thoroughly mixed the data collection forms to insure they were no longer in alphabetical order. The data collection forms were then numbered from 1 to 111 with the identifying number being placed near the space provided for the name of a student. The identifying number was placed on the form a second time in association with the recorded scores.

The final stage of the process designed to disassociate names and scores involved the separation of the student name and identifying number from the identifying number and recorded scores. The portion of the data collection form which included student name and identifying number was placed in a sealed envelope and retained by the elementary school supervisor.

The Metropolitan Achievement Tests: Elementary Form F were given to 153 fourth grade students in October 1977. One hundred eleven of these students continued to be tested in the fifth grade by the Metropolitan Achievement Tests: Intermediate Form G. These 111 students were the studied group.

The design of the study mandated that the 111 selected students be divided into low, average, and high reading ability groups. The

Metropolitan Achievement Tests listed three measures of reading achievement. These were word knowledge, word analysis, and total reading. In order to establish criteria for selection of the reading ability groups the three measures of 1977-78 reading achievement were correlated with measures of science achievement for 1978-79 and 1979-80. Results of the correlations appear in Table II.

TABLE II

Correlations Produced by Comparing 1977-78 Reading Achievement Subpart Scores and Science Achievement Scores for 111 Students in 1978-79 and 1979-80

	Science Achievement	
	1978-79	1979-80
1977-78 Word Knowledge	.783	.697
1977-78 Word Analysis	.828	.757
1977-78 Total Reading	.847	.763

A comparison of the Spearman Rank Correlation coefficients presented in Table II indicated that all coefficients are significant at the 5 percent level of significance. The total reading scores provided the higher correlation coefficients with science achievement for each of the two years. The decision was made to form the three ability groups after consideration of the total reading achievement measures for the 111 students as compiled from scores on the Metropolitan Achievement Tests: Elementary Form F which were administered in October 1977.

The low reading ability group was composed of those students with total reading achievement scores of 4.0 or less on a grade equivalent scale. The average reading ability group included students with total reading achievement scores between 4.1 and 6.0. The high reading ability group included those students with total reading achievement scores of 6.1 or higher. The established intervals contained approximately two grade levels each. The resulting reading ability groups included 27, 50, and 34 students respectively. This study was based on the premise that these groups were statistically different from one another when total reading achievement was considered. Ability group formation can be traced in Figure 2. A compilation of data collected for treatment in this study for the low, average, and high reading ability groups is displayed in Table III, Table IV, and Table V respectively.

IV. READABILITY OF ACHIEVEMENT TESTS

The science achievement data for this study were obtained from the Metropolitan Achievement Tests: Intermediate Form G which were administered at the sixth grade level in April 1980, to the low, average, and high reading ability groups. The major theme for this study was the investigation of the effects of elementary school science readability levels on science achievement.

In order to isolate the effects of elementary school science textbook readability in this study, it was necessary to determine the readability level of the science achievement portion of the Metropolitan

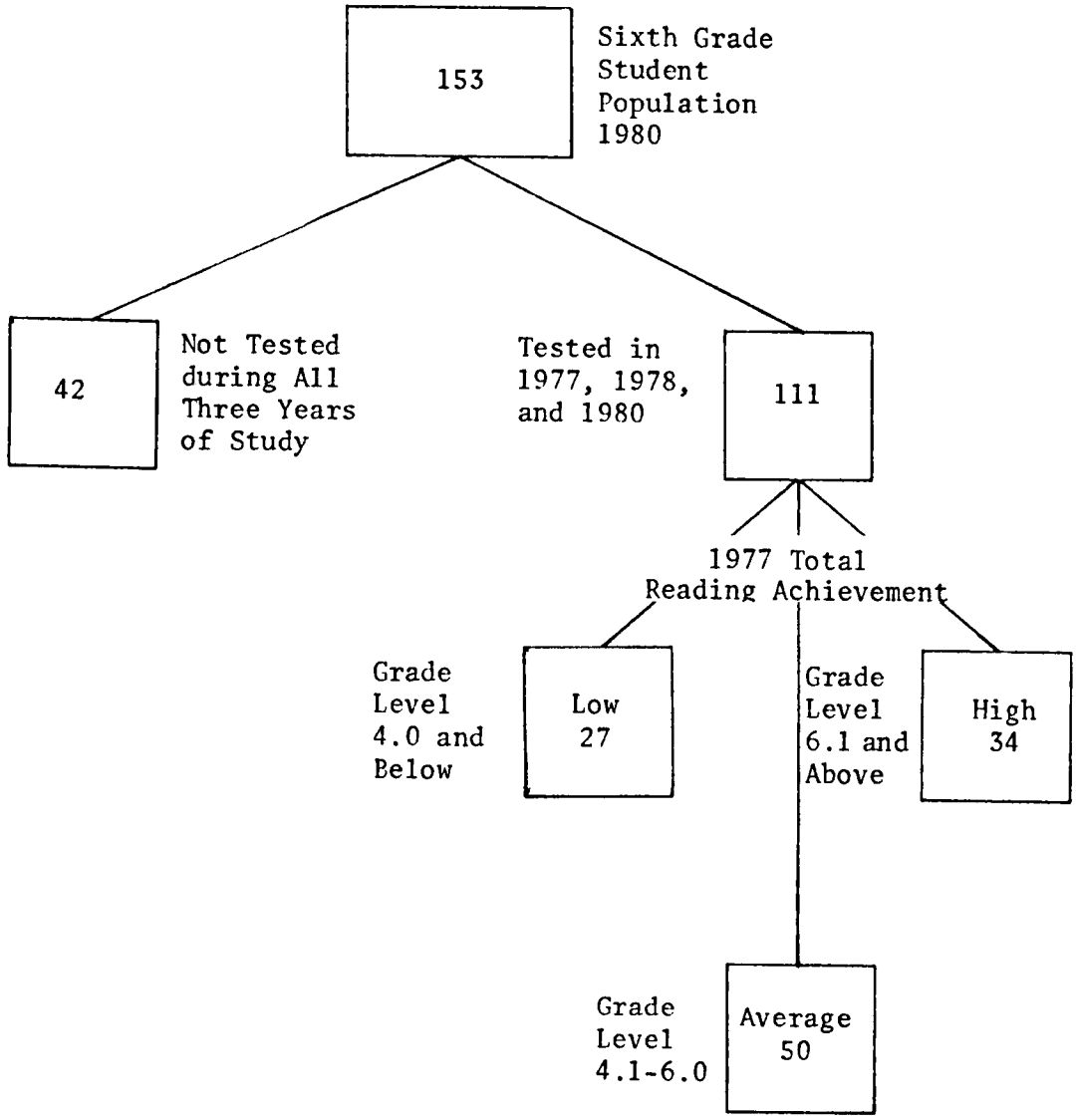


Figure 2. Selection of Low, Average, and High Reading Ability Groups.

TABLE III

Metropolitan Achievement Tests Scores in Grade Level Equivalents for Total Reading Achievement and Science Achievement Compiled for a Low Reading Ability Group during the Period 1977-80

Total Reading 1977-78	Total Reading 1978-79	Science 1978-79	Total Reading 1979-80	Science 1979-80
2.1	2.7	2.7	3.7	3.0
2.2	2.6	3.2	3.9	3.4
2.3	3.5	2.6	4.5	2.6
2.5	3.0	3.4	4.9	4.8
2.8	4.3	3.6	5.5	6.0
2.9	5.6	4.2	4.9	7.1
2.9	4.2	5.3	5.0	4.8
2.9	7.8	4.4	4.8	3.8
3.0	3.4	3.4	4.9	5.4
3.1	3.9	2.8	3.9	3.8
3.2	3.7	4.0	6.0	5.5
3.3	4.4	2.9	4.9	4.2
3.3	4.5	3.3	5.5	6.0
3.3	4.3	4.4	5.6	5.3
3.3	4.4	4.6	5.6	7.8
3.3	4.8	4.6	5.6	4.8
3.3	3.6	3.6	4.5	6.2
3.5	3.4	5.0	4.8	5.4
3.5	4.4	5.5	5.3	6.5
3.6	4.5	4.0	5.6	5.8
3.7	5.3	5.8	6.9	7.0
3.7	3.8	4.2	5.6	5.8
3.7	4.2	2.5	4.6	4.0
3.8	5.0	4.6	6.4	7.0
3.8	4.5	5.3	5.0	5.8
3.9	4.9	6.2	6.4	7.3
3.9	4.1	3.9	5.3	5.0

TABLE IV

Metropolitan Achievement Tests Scores in Grade Level Equivalents for Total Reading Achievement and Science Achievement Compiled for an Average Reading Ability Group during the Period 1977-80

Total Reading 1977-78	Total Reading 1978-79	Science 1978-79	Total Reading 1979-80	Science 1979-80
4.1	4.5	5.3	6.9	9.9
4.1	5.0	5.3	6.7	5.7
4.1	6.0	4.6	9.8	7.4
4.1	4.8	5.3	5.5	5.7
4.2	5.5	5.1	6.6	6.2
4.2	5.0	5.0	6.6	6.5
4.2	5.0	5.7	7.1	6.5
4.2	5.5	4.8	6.6	6.5
4.3	4.8	5.8	6.6	6.2
4.4	6.2	6.2	6.9	7.4
4.4	4.8	5.5	6.6	6.0
4.4	5.3	6.0	6.9	6.5
4.4	5.2	5.5	7.6	7.8
4.4	6.0	6.8	7.8	7.6
4.5	5.3	5.4	6.2	6.5
4.5	5.5	6.0	6.6	9.9
4.6	5.6	5.7	7.4	6.8
4.6	5.0	4.0	6.4	7.1
4.6	5.0	6.0	7.1	7.6
4.8	5.7	5.8	8.0	8.1
4.8	5.5	7.1	6.4	9.9
4.8	4.9	6.2	6.2	7.0
4.8	6.0	6.2	7.6	9.9
4.9	5.6	6.2	6.6	6.2
4.9	5.6	6.0	7.4	7.4
4.9	5.7	6.8	7.4	9.9
4.9	4.8	5.3	6.7	7.3
5.0	6.0	6.2	8.4	7.3
5.0	5.6	7.3	9.9	9.9
5.2	5.8	5.8	6.7	6.5
5.3	5.6	6.8	7.8	8.1
5.3	5.6	5.1	7.8	8.1
5.3	5.8	6.5	8.4	9.9
5.3	5.2	6.0	6.7	8.7
5.5	5.7	5.4	7.6	9.9
5.6	5.8	5.5	8.0	9.9
5.6	5.3	7.0	7.6	7.4
5.6	6.0	5.8	6.9	8.3
5.6	6.0	7.0	9.0	8.0

TABLE IV (Cont'd)

Total Reading 1977-78	Total Reading 1978-79	Science 1978-79	Total Reading 1979-80	Science 1979-80
5.6	6.6	7.1	8.2	8.3
5.6	5.7	6.5	6.9	7.1
5.6	7.1	6.8	8.4	9.9
5.6	6.4	7.4	9.0	9.9
5.6	5.8	6.8	6.7	8.3
5.7	6.4	6.5	7.3	6.8
5.7	6.2	6.8	9.9	9.9
5.7	7.3	7.8	9.9	8.1
6.0	5.8	6.5	9.0	9.9
6.0	6.6	6.8	6.9	8.3
6.0	6.2	6.5	8.0	8.0

TABLE V

Metropolitan Achievement Tests Scores in Grade Level Equivalents for Total Reading Achievement and Science Achievement Compiled for a High Reading Ability Group during the Period 1977-80

Total Reading 1977-78	Total Reading 1978-79	Science 1978-79	Total Reading 1979-80	Science 1979-80
6.2	6.0	5.7	8.0	6.2
6.2	6.7	5.8	6.9	6.8
6.2	7.3	5.8	9.9	8.3
6.2	6.0	6.0	7.3	6.8
6.2	6.7	7.0	7.3	8.0
6.4	7.3	9.9	9.9	9.9
6.4	7.1	7.8	9.9	9.4
6.4	6.7	7.6	9.9	8.0
6.4	6.2	5.3	7.3	7.3
6.6	6.7	7.3	8.0	7.6
6.6	6.4	7.3	6.6	8.0
6.9	7.1	7.1	9.8	9.9
6.9	6.2	6.5	7.1	6.8
7.3	6.7	7.1	9.9	8.0
7.3	8.0	9.9	9.9	9.4
7.3	7.8	8.3	9.9	9.9
7.3	8.2	7.3	9.9	9.9
7.6	7.1	6.2	9.9	7.8
7.6	7.8	6.8	7.3	8.7
7.6	6.7	6.8	8.0	9.4
7.6	9.9	7.4	9.9	9.4
7.6	7.6	5.5	9.9	9.9
8.0	7.1	7.3	9.9	8.7
8.0	9.9	8.7	9.9	9.9
8.0	9.9	7.6	9.9	9.9
8.0	8.7	8.1	9.9	9.9
8.0	8.2	9.9	9.9	9.9
8.0	9.8	8.7	9.9	9.9
8.0	9.9	9.9	9.9	9.9
8.0	9.9	9.9	9.9	9.9
8.0	8.2	8.0	9.9	9.9
8.0	8.4	9.9	9.9	9.9
8.0	7.3	6.8	8.0	9.9
8.0	7.1	6.8	9.9	9.9

Achievement Tests: Intermediate Form G. Three samples were selected for analysis which resembled textbook material in format. Each sample was composed of 100 words and the samples were selected from different areas of the science achievement portion. When multiple choice responses were included in a sample, the most difficult response to read was analyzed.

Consistency of readability measures was preserved by employing the Fry Readability Graph to determine the estimated readability for each sample. The estimated readability level determined for the science achievement portion of the Metropolitan Achievement Tests: Form G was sixth grade level.

V. SELECTION OF STATISTICAL PROCEDURES

Treatment of the data collected for this study required both independent group and related group procedures. Data obtained to compare groups on reading achievement and science achievement required independent group procedures. Data obtained to compare gains on reading achievement with gains on science achievement within the same group mandated the use of related group procedures.

The statistical options available for implementation in treating the data were of either a general or a specific nature. General statistical procedures would have provided answers regarding differences among the ability groups without providing specific answers regarding

any two ability groups.[4] The design of the study mandated specific answers regarding differences between any two ability groups, if such differences existed. Procedures such as the t test for independent groups, Mann-Whitney U test, t test for paired data, and Wilcoxon Matched-Pairs Signed-Ranks test were then considered for use where appropriate.

Consideration of the aforementioned statistical procedures necessitated a choice between parametric and nonparametric procedures. The data collected for this study were recorded in grade equivalents. Assumptions associated with parametric statistical procedures included the assumption that the data are at least interval in scale.[5] Although the grade equivalent scale is interval in format, this characteristic is lost as associations with the original raw scores are made. The grade equivalents used in this study were categorized as ordinal in scale. The nonparametric procedures were the statistical procedures selected.

Based on the discussion above, the statistical procedures selected were nonparametric procedures which provided specific answers as to differences between ability groups or between two sets of related data. Data obtained from independent groups were treated with the Mann-Whitney U test. Data obtained from related groups were treated with the

---

[4]Roger E. Kirk, Experimental Design: Procedures for the Behavioral Sciences, (Belmont, California: Brooks/Cole Publishing Company, 1968), p. 73.

[5]Vicki F. Sharp, Statistics for the Social Sciences, (Boston: Little, Brown and Company, 1979), pp. 266-267.

Wilcoxon Matched-Paris Signed-Ranks test. Nominal data were treated by the 2 × 2 contingency table and/or Fisher's Exact Probability test where appropriate.

Mann-Whitney U Test

The Mann-Whitney U test was deemed appropriate for use in this study based on rationale provided by Siegel. According to Siegel:

> When at least ordinal measurement has been achieved, the Mann-Whitney U test may be used to test whether two independent groups have been drawn from the same population. This is one of the most powerful of the nonparametric tests, and it is a most useful alternative to the parametric "t" test when the researcher wishes to avoid the "t" test's assumptions, or when the measurement in the research is weaker than interval scale.[6]

Group size and frequency of tied scores dictated the use of a formula version intended for groups larger than twenty individuals with a correction process to compensate for tied scores. The formula selected was:

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\left(\frac{n_1 n_2}{N(N-1)}\right)\left(\frac{N^3 - N}{12}\right) - \Sigma T}}$$

when U is computed by:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

---

[6]Sidney Siegel, Nonparametric Statistics for the Behavioral Sciences, (New York: McGraw-Hill Book Company, 1956), p. 116.

The ΣT is computed by summing the values obtained from computing $\frac{t^3-t}{12}$ for each of the tied observations, when t is the number of times a particular tie occurs.[7]

The z values arrived at by use of the Mann-Whitney U test were compared to z values obtained from a Table of z Values. The tabled z values selected for comparison reflected the desired significance level as mandated by the hypothesis under consideration.

The following hypotheses were tested using the Mann-Whitney U test:

1. There is no significant difference between a low reading ability group and an average reading ability group as compared on total reading achievement scores measured by the Metropolitan Achievement Tests: Elementary Form F administered in October 1977.

2. There is no significant difference between an average reading ability group and a high reading ability group as compared on total reading achievement scores measured by the Metropolitan Achievement Tests: Elementary Form F administered in October 1977.

3. There is no significant difference between a low reading ability group and an average reading ability group as compared on science achievement scores measured by the Metropolitan Achievement Tests: Intermediate Form G administered in April 1980.

4. There is no significant difference between an average reading ability group and a high reading ability group as compared on science

---

[7]Ibid., pp. 123-125.

achievement scores measured by the Metropolitan Achievement Tests: Intermediate Form G administered in April 1980.

Hypotheses directed at total reading achievement were compared at the .05 level of significance while the region of rejection was one-tailed. Hypotheses involving science achievement were compared at the .05 level of significance with the region of rejection being two-tailed.

Wilcoxon Matched-Pairs Signed-Ranks Test

The Wilcoxon Matched-Pairs Signed-Ranks test was selected for use in this study. Several other nonparametric procedures were available, but all lacked the combination of direction of differences and magnitude of direction of differences provided by the Wilcoxon Matched-Pairs Signed-Ranks test. With regard to the last characteristic, which is not included in the other procedures, Siegel writes:

> . . . The Wilcoxon Matched-Pairs Signed-Ranks test does just that: it gives more weight to a pair which shows a large difference between the two conditions than to a pair which shows a small difference.[8]

Due to the size of the ability groups, the format for large samples was chosen to treat the data. The formula used was:

$$z = \frac{T - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)\ (2N+1)}{24}}}$$

The T is the sum of the ranks of the less frequent sign and N is the count of differences between observations other than zero.[9]

---

[8]Ibid., p. 75.

[9]Ibid., pp. 79-82.

As with the z values computed with the Mann-Whitney U test, the z values computed with the Wilcoxon Matched-Pairs Signed-Ranks test are compared to a tabled z value. The tabled z value reflected the selected level of significance for the hypothesis under consideration.

The following hypotheses were tested using the Wilcoxon Matched-Pairs Signed-Ranks test:

5. There is no significant difference between total reading achievement gains and science achievement gains for a low reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

6. There is no significant difference between total reading achievement gains and science achievement gains for an average reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

7. There is no significant difference between total reading achievement gains and science achievement gains for a high reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

Hypotheses which concentrated on gains in total reading achievement and science achievement were tested at the .05 significance level. The region of rejection was set in two-tails of the curve.

2 × 2 Contingency Table

The 2 × 2 contingency table was selected to make comparisons where the data were presented in nominal form. In order to compensate for cells with small entries, the procedures for the 2 × 2 contingency table which incorporated the Yate's correction for continuity was chosen. The formula used was:

$$\chi^2 = \frac{(|ad - bc| - n/2)^2 \, n}{(a+b)(a+c)(b+d)(d+c)}$$

when degrees of freedoms were computed by:

$$df = (c-1)(r-1)$$ [10]

The computed chi square values were compared to a tabled chi square value of 3.841. This comparison was appropriate since all 2 × 2 contingency tables have two columns and two rows producing degrees of freedom equal to 1.

The following hypotheses were tested using the 2 × 2 contingency table:

8. There is no significant difference between low reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

9. There is no significant difference between average reading ability students with total reading achievement scores of 6.0 or above

---

[10]Helen M. Walker and Joseph Lev, Elementary Statistical Methods (New York: Holt, Rinehart and Winston, Inc., 1969), pp. 280-282.

and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

10. There is no significant difference between high reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

11. There is no significant difference between low reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

12. There is no significant difference between average reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

13. There is no significant difference between high reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

Fisher's Exact Probability Test

The design of this study made possible entries in cells which would be too small to be compensated for by the Yates' correction for continuity. When this condition occurred the appropriate statistical procedure to be used was the Fisher's Exact Probability test.[11]

---

[11] Ibid., p. 82.

Siegel suggested the following formula for arriving at p by Fisher's Exact Probability test:

$$p = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{N!A!B!C!D!}$$

The computed p was multiplied by 2 before comparing to a standard of .05 which provided a two-tailed region of rejection with a .05 level of significance.[12]

[12]Siegel, pp. 96-100.

CHAPTER IV

FINDINGS

The purpose of this study was to investigate some of the effects of elementary school science textbook readability levels on science achievement of elementary school students with low, average, and high reading abilities as measured by the Metropolitan Achievement Tests: Science. To this end thirteen hypotheses were tested.

Data collected for study were obtained from students who had studied a basal elementary school science textbook series during the fifth and sixth grades which was written above grade level. In addition they had been regularly tested using the Metropolitan Achievement Tests in the fourth through the sixth grade levels.

At the fourth grade level the Metropolitan Achievement Tests: Elementary Form F were administered and became the source of reading achievement scores which were used to form the reading ability groups. The Metropolitan Achievement Tests: Intermediate Form were administered at the fifth and sixth grade levels with forms F and G administered respectively. The latter forms of the Metropolitan Achievement Tests provided both reading achievement and science achievement scores.

In October of 1977, 153 fourth grade students were administered the Metropolitan Achievement Tests: Elementary Form F. Of the original 153 fourth grade students, 111 were later to be tested with the Metropolitan Achievement Tests: Intermediate Form at the fifth and sixth

grade levels. These 111 students were designated as the study group. From this study group three smaller groups were selected based upon reading achievement scores obtained at the fourth grade level. All forms of the Metropolitan Achievement Tests yield three reading achievement scores. These are word knowledge, word analysis, and total reading. Total reading achievement scores produced higher correlation coefficients with science achievement scores and were used in this study as a basis for reading ability group formation.

Those students who scored 4.0 or less on total reading achievement were designated as the low ability reading group. This group contained 27 students. The average reading ability group contained 50 students who scored between 4.1 and 6.0 in total reading achievement. The high reading ability group included 34 students who scored above 6.1 in total reading achievement.

I. ANALYSIS OF FOURTH GRADE TOTAL READING ACHIEVEMENT

This study was dependent, in part, upon making sixth grade science achievement comparisons among three groups which were significantly different when compared in total reading achievement at the fourth grade level.

Hypothesis I. There is no significant difference between a low reading ability group and an average reading ability group as compared on total reading achievement scores measured by the Metropolitan Achievement Tests: Elementary Form F administered in October 1977.

The Mann-Whitney U test was used to test this hypothesis to determine if there was a significant difference between the low reading ability group and the average reading ability group on fourth grade total reading achievement scores. The computed z was found to be 7.22 with the ΣT equal to 125. The computed z was compared to a tabled z of 1.645 which provided a one-tailed region of rejection with a .05 level of significance. Hypothesis I was rejected.

Hypothesis II. There is no significant difference between an average reading ability group and a high reading ability group as compared in total reading achievement scores measured by the Metropolitan Achievement Tests: Elementary Form F administered in October 1977.

The Mann-Whitney U test was used to test this hypothesis to determine if there was a significant difference between the average reading ability group and the high reading ability group on fourth grade total reading achievement scores. The computed z was found to be 7.77 with the ΣT equal to 276. The computed z was compared to a tabled z of 1.645 which provided a one-tailed region of rejection with a .05 level of significance. Hypothesis II was rejected.

Consideration of the actions taken on these two hypothesis led to the conclusion that the three reading ability groups were different when compared on total reading achievement scores. Information pertinent to the testing of Hypothesis I and Hypothesis II can be found in Table VI.

TABLE VI

Mann-Whitney U Test Analysis of Fourth Grade Total Reading Achievement Scores of Reading Ability Groups

Hypothesis	Ability Group	n	Median	U	ΣT	z
I	Low	27	3.30	1350	125	7.22*
	Average	50	4.90			
II	Average	50	4.90	1700	276	7.77*
	High	34	7.45			

*Significant at the .05 significance level, one-tailed region of rejection.

II. ANALYSIS OF SIXTH GRADE SCIENCE ACHIEVEMENT

Students participating in this study had participated in a science program at the fifth and sixth grade levels which had been based on elementary school science textbooks written above grade level. Comparisons were made among the reading ability groups with regard to sixth grade science achievement.

Hypothesis III. There is no significant difference between a low reading ability group and an average reading ability group as compared on science achievement scores measured by the Metropolitan Achievement Tests: Intermediate Form G administered in April 1980.

The Mann-Whitney U test was used to test this hypothesis to determine if there was a significant difference between the low reading ability group and the average reading ability group on sixth grade science achievement. The computed z was found to be 5.99 with the ΣT

equal to 63.5. The computed z was compared to a tabled z of 1.96 which provided a two-tailed region of rejection with a .05 level of significance. Hypothesis III was rejected.

Hypothesis IV. There is no significant difference between an average reading ability group and a high reading ability group as compared on science achievement scores measured by the Metropolitan Achievement Tests: Intermediate Form G administered in April 1980.

The Mann-Whitney U test was used to test this hypothesis to determine if there was a significant difference between the average reading ability group and the high reading ability group on sixth grade science achievement scores. The computed z was found to be 2.92 with ΣT equal to 82.5. The computed z was compared to a tabled z of 1.96 which provided a two-tailed region of rejection with a .05 level of significance. Hypothesis IV was rejected.

Consideration of the action taken on these two hypotheses led to the conclusion that the three reading ability groups were different when compared on science achievement scores. Information pertinent to the testing of Hypothesis III and Hypothesis IV can be found in Table VII.

III. ANALYSIS OF TOTAL READING ACHIEVEMENT AND SCIENCE ACHIEVEMENT GAINS

In an effort to gain a greater understanding of the relationship between total reading achievement and science achievement, gains in each of these areas were compared. Data used to determine the gains were obtained from the Metropolitan Achievement Tests: Intermediate

TABLE VII

Mann-Whitney U Test Analysis of Sixth Grade Science Achievement Scores of Reading Ability Groups

Hypothesis	Ability Group	N	Median	U	ΣT	z
III	Low	27	5.4	1237	63.5	5.99*
	Average	50	7.7			
IV	Average	50	7.7	1170	82.5	2.92*
	High	34	9.4			

*Significant at the .05 significance level, two-tailed region of rejection.

Form F given in October 1980 and the Metropolitan Achievement Tests: Form G given in April 1980.

Hypothesis V. There is no significant difference between total reading achievement gains and science achievement gains for a low reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

The Wilcoxon Matched-Paris Signed-Ranks test was used to test this hypothesis to determine if there was a significant difference between the total reading achievement gains and the science achievement gains recorded for the low reading ability group. The computed z was found to be 1.14 with T equal to 220.5. The computed z was compared to a tabled z of 1.96 which provided a two-tailed region of rejection with a .05 level of significance. Hypothesis V was not rejected.

Hypothesis VI. There is no significant difference between total reading achievement gains and science achievement gains for an average reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

The Wilcoxon Matched-Pairs Signed-Ranks test was used to test this hypothesis to determine if there was a significant difference between the total reading achievement gains and the science achievement gains recorded for the average reading ability group. The computed z was found to be .12 with T equal to 576.5. The computed z was compared

to a tabled z of 1.96 which provided a two-tailed region of rejection with a .05 level of significance. Hypothesis VI was not rejected.

Hypothesis VII. There is no significant difference between total reading achievement gains and science achievement gains for a high reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

The Wilcoxon Matched-Pairs Signed-Ranks test was used to test this hypothesis to determine if there was a significant difference between the total reading achievement gains and the science achievement gains recorded for the high reading ability group. The computed z was found to be .26 with T equal to 278. The computed z was compared to a tabled z of 1.96 which provided a two-tailed region of rejection with a .05 level of significance. Hypothesis VII was not rejected.

Consideration of the action taken on these three hypotheses led to the conclusion that total reading achievement gains were not significantly different from science achievement gains for any of the three reading ability groups. Information pertinent to the testing of Hypothesis V, Hypothesis VI, and Hypothesis VII can be found in Table VIII.

IV. ANALYSIS OF THE EFFECTS OF READABILITY ESTIMATES ON SCIENCE ACHIEVEMENT

Previous sections of this chapter presented analyses which explored relationships between total reading achievement and science achievement. Readability data were added as a final variable in the analysis included in this section. Data previously analyzed were achievement scores

TABLE VIII

Wilcoxon Matched-Pairs Signed-Ranks Test Analysis of Total Reading Achievement Gains and Science Achievement Gains of Reading Ability Groups

Hypotheses	Ability Groups	n	Corrected n	Median Gains Reading	Median Gains Science	z
V	Low	27	26	1.1	1.2	1.14*
VI	Average	50	48	1.8	1.5	.12*
VII	High	34	32	1.4	1.3	.26*

*Not significant at the .05 significance level, two-tailed region of rejection.

obtained from various forms of the Metropolitan Achievement Tests. The common source of science information for participants of the low, average, and high ability reading groups were elementary school science textbooks.

Readability estimates for the Metropolitan Achievement Tests: Intermediate Form G were the focal points for Hypothesis VIII, Hypothesis IX, and Hypothesis X. Readability estimates of the sixth grade elementary school science textbooks were the focal points for Hypothesis XI, Hypothesis XII, and Hypothesis XIII.

Hypothesis VIII. There is no significant difference between low reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

The Fisher's Exact Probability test was used to test this hypothesis to determine if there was a significant difference between the ability to read the science portion of the Metropolitan Achievement Tests: Intermediate Form G and the attainment of grade level science achievement for a low ability reading group. The computed p was found to be .013 and became .026 when multiplied by two. The computed p of .026 was compared to a standard of .05 which provided a two-tailed region of rejection with a .05 level of significance. Hypothesis XIII was rejected.

Hypothesis IX. There is no significant difference between average reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

The Fisher's Exact Probability test was used to test this hypothesis to determine if there was a significant difference between the ability to read the science portion of the Metropolitan Achievement Tests: Intermediate Form G and the attainment of grade level science achievement for an average ability reading group. The computed p was found to be .240 and became .480 when multiplied by two. The computed p of .480 was compared to a standard of .05 which provided a two-tailed region of rejection with a .05 level of significance. Hypothesis IX was not rejected.

Hypothesis X. There is no significant difference between high reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

The Fisher's Exact Probability test was used to test this hypothesis to determine if there was a significant difference between the ability to read the science portion of the Metropolitan Achievement Tests: Intermediate Form G and the attainment of grade level science achievement for a high ability reading group. The computed p was found to be 1.000 and became 2.000 when multiplied by two. The computed p of 2.000 was compared to a standard of .05 which provided a two-tailed region of rejection with a .05 level of significance. Hypothesis X was not rejected.

Hypothesis XI. There is no significant difference between low reading ability students with total reading achievement scores of 7.0

or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

The Fisher's Exact Probability test was used to test this hypothesis to determine if there was a significant difference between the ability to read the sixth grade science textbook and the attainment of grade level science achievement for a low ability reading group. The computed p was found to be 1.000 and became 2.000 when multiplied by two. The computed p of 2.000 was compared to a standard of .05 which provided a two-tailed region of rejection with a .05 level of significance. Hypothesis XI was not rejected.

Hypothesis XII. There is no significant difference between average reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

The Fisher's Exact Probability test was used to test this hypothesis to determine if there was a significant difference between the ability to read the sixth grade science textbook and the attainment of grade level science achievement for an average ability reading group. The computed p was found to be .0003 and became .0006 when multiplied by two. The computed p of .0006 was compared to a standard of .05 which provided a two-tailed region of rejection with a .05 level of significance. Hypothesis XII was rejected.

Hypothesis XIII. There is no significant difference between high reading ability students with total reading achievement scores of 7.0 or

above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

The Fisher's Exact Probability test was used to test this hypothesis to determine if there was a significant difference between the ability to read the sixth grade science textbook and the attainment of grade level science achievement for a high ability reading group. The computed p was found to be .940 and became 1.880 when multiplied by two. The computed p of 1.880 was compared to a standard of .05 which provided a two-tailed region of rejection with a .05 level of significance. Hypothesis XIII was not rejected.

The Fisher's Exact Probability test was designed to determine whether two groups differ in the proportion with which they fall into two classifications. It is normally employed when extremely small cell frequencies are observed. Hypothesis IX, Hypothesis X, Hypothesis XI, and Hypothesis XIII were not rejected when treated by the Fisher's Exact Probability test. Examination of the data treated by these hypotheses indicated small cell frequencies and a disproportionate representation for one of the two groups being analyzed.

Hypothesis IX and Hypothesis X centered on ability to read the science portion of the Metropolitan Achievement Tests: Intermediate Form G. The high total reading achievement scores of both the average and high reading ability groups severely limited participation in the group composed of the below 6.0 on total reading achievement. Further analysis of this data also indicated that the majority of the students with total reading achievement scores above the 6.0 level also had science achievement scores above 6.7.

Hypothesis XI and Hypothesis XIII dealt with ability to read the sixth grade science textbook. Hypothesis XI required the treatment of data relevant to the low reading ability group. Review of this data indicated that few students scored above the 7.0 level on total reading achievement. This group also tended to score below 6.7 on science achievement. The high reading ability group was composed of a majority of students who scored above 7.0 on total reading achievement and above 6.7 on science achievement. This data was addressed through Hypothesis XIII.

Hypothesis VIII was rejected. Data analyzed through this hypothesis dealt with the ability of a low reading ability group to read the science portion of the Metropolitan Achievement Tests: Intermediate Form G and to score above 6.7 on science achievement. Analysis of the data indicated that there was a tendency to score above 6.7 on science achievement when total reading achievement was 6.0 or higher. Those students scoring below 6.0 on total reading achievement tended to score below 6.7 on science achievement.

Hypothesis XII was rejected. Data analyzed through this hypothesis pertained to an average reading ability group and was concerned with the ability to read sixth grade science textbooks and to score above 6.7 on science achievement. Analysis of this data indicated that there was a strong tendency for students to score above 6.7 on science achievement when they scored above 7.0 on total reading achievement. Those students who scored below 7.0 on total reading achievement tended to have an equal opportunity to score above or below 6.7 on science achievement.

Presentation of the results for Hypothesis VIII through Hypothesis XIII can be found in Table IX. Presentation of data treated by these hypotheses can be found in Figure 3, Figure 4, and Figure 5.

TABLE IX

Fisher's Exact Probability Test Analysis of the Effects of Readability Estimates for the Metropolitan Achievement Tests: Science and the Sixth Grade Science Textbook on Grade Level Science Achievement

Hypotheses	Ability Group	n	Computed p	
			MAT:Science	Textbook
VIII and XI	Low	27	.026*	2.000
IX and XII	Average	50	.480	.0006*
X and XIII	High	34	2.000	1.880

*Significant at the .05 significance level, two-tailed region of rejection.

V. ANALYSIS OF DATA BY INSPECTION OF INDIVIDUAL READING ABILITY GROUPS AND THE TOTAL READING ABILITY GROUP

The Metropolitan Achievement Tests: Intermediate Form G were administered in April 1980, at the 6.7 grade level. In this study those students who scored on or above 6.7 were considered to have attained grade level science achievement. Hypothesis VIII, Hypothesis IX, and Hypothesis X were designed to determine if those students who scored less than 6.0 on total reading achievement also scored less than 6.7 on science achievement and if those students who scored on

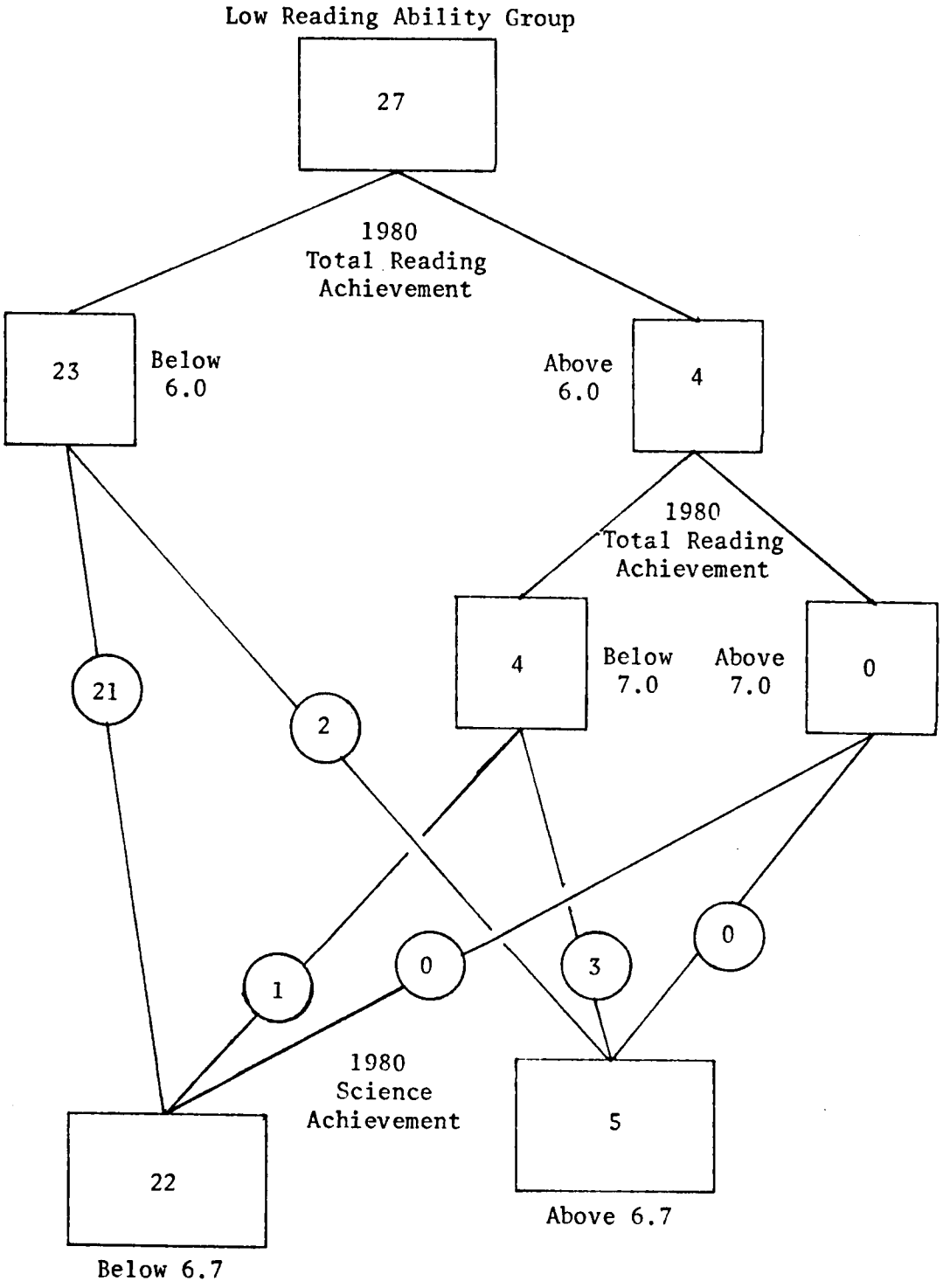


Figure 3. Presentation of the Effect of Total Reading Achievement on Grade Level Science Achievement for a Low Reading Ability Group.

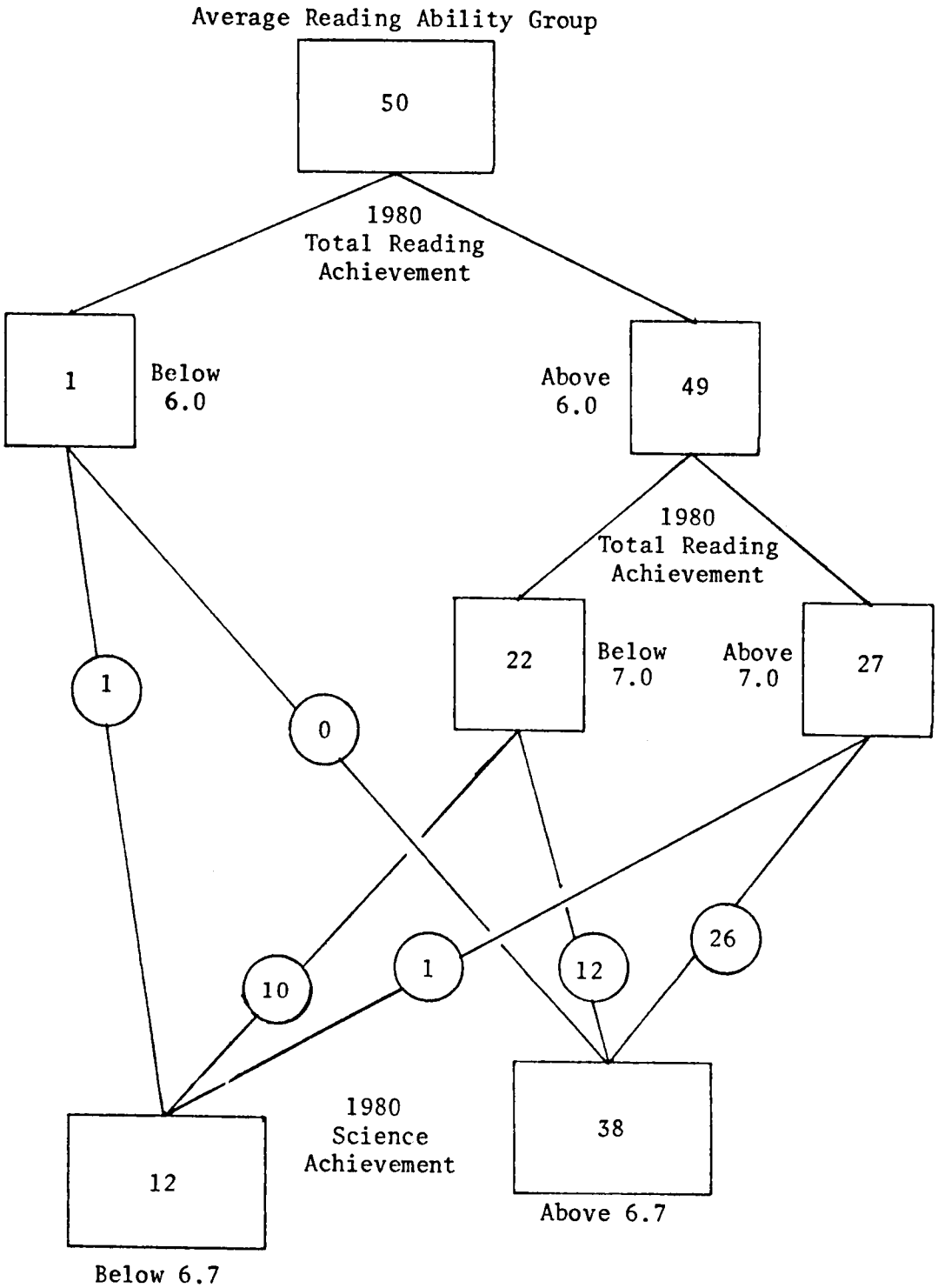


Figure 4. Presentation of the Effect of Total Reading Achievement on Grade Level Science Achievement for an Average Reading Ability Group.

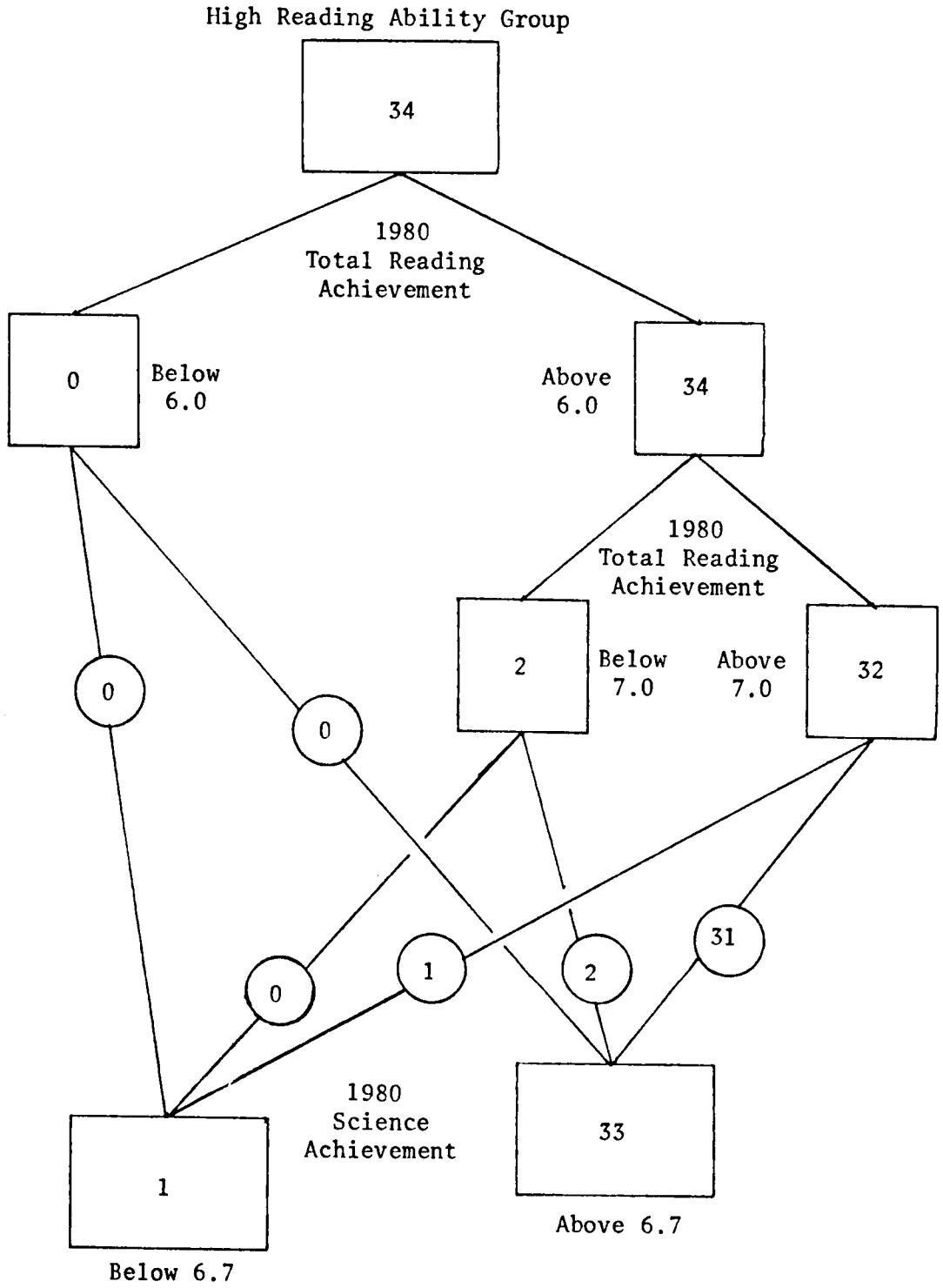


Figure 5. Presentation of the Effect of Total Reading Achievement on Grade Level Science Achievement for a High Reading Ability Group.

or above 6.0 on total reading achievement also scored on or above 6.7 on science achievement. The science portion of this achievement test had an estimated readability of 6.0.

The sixth grade science textbook was estimated to have a readability of 7.0. Hypothesis XI, Hypothesis XII, and Hypothesis XIII were designed to see if those students who scored less than 7.0 on total reading achievement also scored less than 6.7 on science achievement and if those students who scored on or above 7.0 on total reading achievement scored on or above 6.7 on science achievement. Table X includes data relevant to those students who did not perform as expected. Due to the relationship of the two reading comparisons with science achievement, data relevant to all six hypotheses was included in Table X.

Analysis of the data in Table X indicates that there were 21 students of the 111 students in the study who performed in ways which were not expected with regard to comparisons of ability to read the sixth grade science textbook and science achievement scores. Of the 21 students only two did not attain grade level science achievement when they were expected to, based on the reading achievement criteria. In both cases the students were expected to be able to read the Metropolitan Achievement Tests: Intermediate Form G, and were expected to read the sixth grade science textbook.

All of the remaining 19 students attained grade level science achievement scores. Only two of the 19 students were expected to experience difficulty with reading the science portion of the achievement

TABLE X

Individual Analysis of Data by Group for Comparisons of Total Reading Achievement to 6.0 and 7.0 with Science Achievement Scores Obtained from the Metropolitan Achievement Tests: Intermediate Form G

	Total Reading Achievement			Science Achievement		
Group	Total Reading Score	Relation to 6.0 (MAT)	Relation to 7.0 (Textbook)	Science Score	Expected Relation to 6.7	Observed Relation to 6.7
Low	5.6	B	B	7.8	B	A
	4.9	B	B	7.1	B	A
	6.9	A	B	7.0	B	A
	6.4	A	B	7.3	B	A
	6.4	A	B	7.0	B	A
Average	6.9	A	B	9.9	B	A
	6.9	A	B	8.3	B	A
	6.9	A	B	8.3	B	A
	6.9	A	B	7.4	B	A
	6.9	A	B	7.1	B	A
	6.7	A	B	8.7	B	A
	6.7	A	B	8.3	B	A
	6.7	A	B	8.3	B	A
	6.6	A	B	9.9	B	A
	6.4	A	B	9.9	B	A
	6.4	A	B	7.1	B	A
	6.2	A	B	7.0	B	A
	7.1	A	A	6.5	A	B
High	6.9	A	B	6.8	B	A
	6.6	A	B	8.0	B	A
	8.0	A	A	6.2	A	B

A = Above Criterion B = Below Criterion

test and the sixth grade science textbook. The other 17 students all were expected to read the science portion of the achievement test and were not expected to read with ease the sixth grade science textbook.

The majority of the students not performing in the expected ways were included in the average reading ability group or the low reading ability group. Together the two groups included 18 of the 21 students who did not perform in the expected ways. All but two of the 18 students attained a total reading achievement score within one grade level of 7.0. This was the estimated readability level of the sixth grade science textbook.

There is an additional note with regard to students in the average reading ability group who did not attain a sufficiently high science achievement score to be considered at grade level. All ten of the students scored within one grade level of 7.0 in total reading achievement. This group occupied a similar range to 12 other students who were expected to read the science portion of the achievement test and not the sixth grade science textbook. The latter group scored above grade level in science achievement. Source of this information is Table IV, page 79, and Figure 4, page 107.

In reference to Table III, page 78, Table IV, and Table V, page 81, 97 percent of the students from all groups with total reading achievement scores on or above 6.0 attained grade level science achievement scores of 6.7 or greater. Only 9 percent of those with total reading achievement scores less than 6.0 attained grade level science achievement. With reference to a total reading achievement score of 7.0 or greater, 96 percent of those students attained grade level science achievement.

Those students with a total reading achievement score of less than 7.0 attained grade level science achievement at a 57 percent rate. This same information may be derived from a study of Figure 6 which presents summary data for all reading ability groups. Data in Figure 6 was previously presented by reading ability group in Figure 3, page 106, Figure 4, page 107, and Figure 5, page 108.

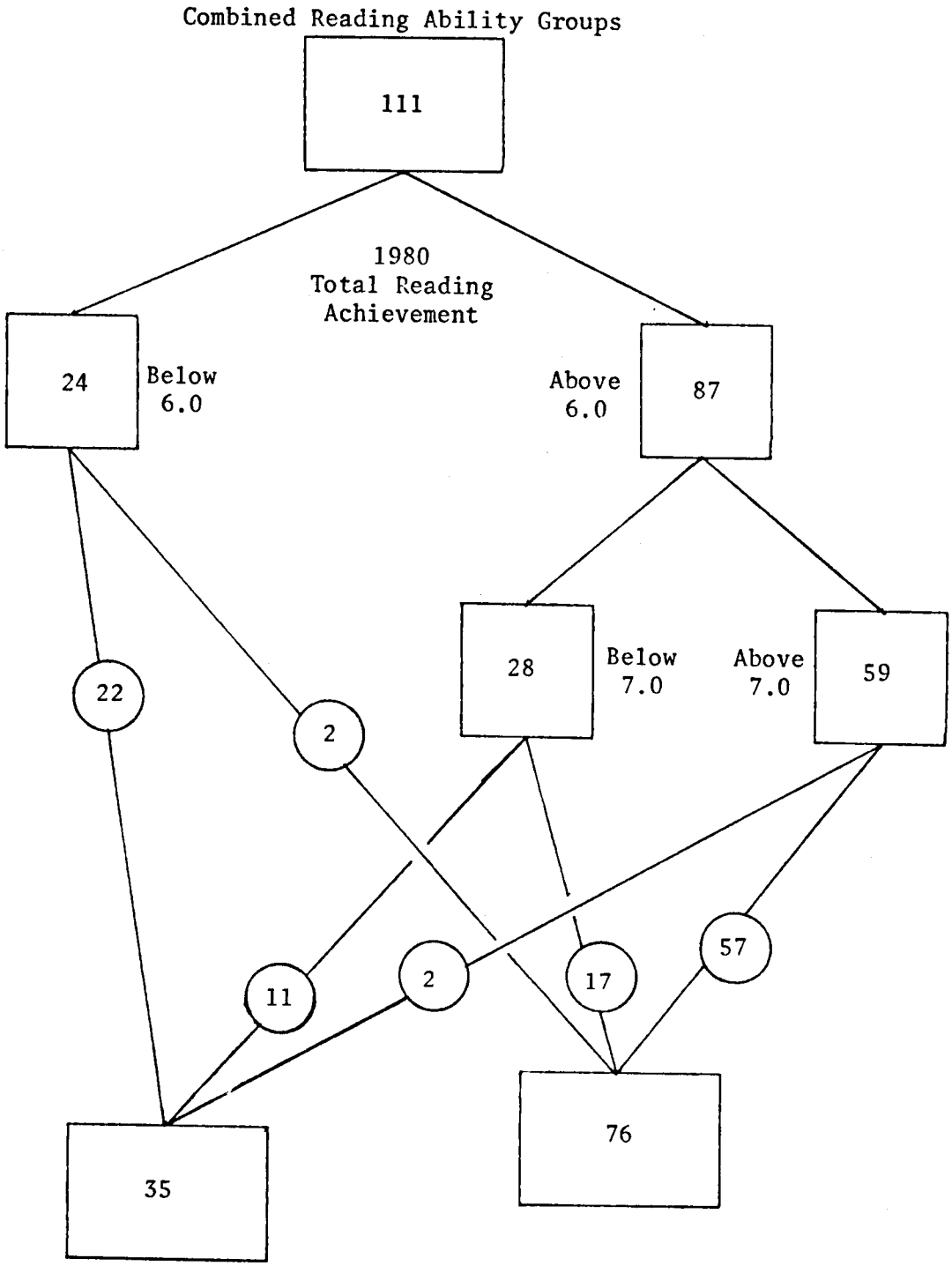


Figure 6. Presentation of the Combined Reading Ability Groups.

CHAPTER V

SUMMARY AND CONCLUSIONS

I. SUMMARY

The purpose of this study was to investigate some of the effects of elementary science textbook readability levels on science achievement of elementary school students with low, average, and high reading abilities as measured by the Metropolitan Achievement Tests: Science.

This study involved 111 students selected from a population of 153 sixth grade students. These students had been annually tested with the Metropolitan Achievement Tests and had studied science from a basal science textbook series which included textbooks that possessed estimated readabilities above the grade level for which they were intended.

The 111 students selected for this study had been tested at the fourth grade level with the Metropolitan Achievement Tests: Elementary Form F administered in October 1977. They had also been tested at the fifth and sixth grade levels with the Metropolitan Achievement Tests: Intermediate Form F and G respectively in October 1978, and April 1980.

The total reading achievement score obtained on the Metropolitan Achievement Tests: Elementary Form F was used to subdivide the 111 students into three reading ability groups. The low reading ability group was composed of those students with total reading achievement scores of 4.0 or less. The average reading ability group had total

reading achievement scores of 4.1 to 6.0. The high reading ability group possessed total reading achievement scores of 6.1 and above. The three groups contained 27, 50, and 34 members respectively.

In order to establish that the reading ability groups were significantly different when compared in total reading achievement, the Mann-Whitney U test was employed. Using a one-tail region of rejection and a .05 level of significance the following hypotheses were tested:

1. There is no significant difference between a low reading ability group and an average reading ability group as compared on total reading achievement scores measured by the Metropolitan Achievement Tests: Elementary Form F administered in October 1977.

2. There is no significant difference between an average reading ability group and a high reading ability group as compared on total reading achievement scores measured by the Metropolitan Achievement Tests: Elementary Form F administered in October 1977.

Science achievement data obtained from the Metropolitan Achievement Tests: Intermediate Form G administered in April 1980 were compared by the Mann-Whitney U test. A two-tailed region of rejection was selected with a .05 level of significance. The following hypotheses were tested:

3. There is no significant difference between a low reading ability group and an average reading ability group as compared on science achievement scores measured by the Metropolitan Achievement Tests: Intermediate Form G administered in April 1980.

4. There is no significant difference between the average reading ability group and the high reading ability group as compared on science

achievement scores measured by the Metropolitan Achievement Tests: Intermediate Form F administered in April 1980.

Total reading achievement scores and science achievement scores were taken from the Metropolitan Achievement Tests: Intermediate Forms F and G which had been administered respectively in October 1978 and April 1980. Gains in total reading achievement and science achievement were computed. These gains were than subjected to test by the Wilcoxon Matched-Pairs Signed-Rank test. A two-tail region of rejection with a .05 level of significance was selected. The following hypotheses were tested:

5. There is no significant difference between total reading achievement gains and science achievement gains for a low reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

6. There is no significant difference between total reading achievement gains and science achievement gains for an average reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

7. There is no significant difference between total reading achievement gains and science achievement gains for a high reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

In order to determine the relation between readability of materials and the attainment of grade level science achievement the

Fisher's Exact Probability test was selected. Attempts were made to use both the 2 × 2 contingency table and the 2 × 2 contingency table with the Yate's correction for continuity. Since neither of these were appropriate due to small cell entries, the Fisher's Exact Probability test was selected. The tests were dependent on making comparisons of total reading achievement with the estimated readabilities of the science portion of the Metropolitan Achievement Tests: Intermediate Form G and the sixth grade elementary school science textbook. The estimated readabilities were 6.0 for the former and 7.0 for the latter. The aforementioned reading comparisons were made in connection with the attainment of grade level science achievement which was deemed to be 6.7. The region of rejection was two-tailed with the level of significance set at .05. The following hypotheses were tested:

8. There is no significant difference between low reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

9. There is no significant difference between average reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

10. There is no significant difference between high reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

11. There is no significant difference between low reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

12. There is no significant difference between average reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

13. There is no significant difference between high reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

Data included in the testing of the six previous hypotheses were subjected to visual comparison. Of particular interest were the low reading ability group and the average reading ability group. Total reading achievement scores between 6.0 and 7.0 were of greatest interest.

II. CONCLUSIONS

A review of the findings led to the following actions being taken on the hypotheses under consideration during this study.

Hypothesis I. There was a significant difference between a low reading ability group and an average reading ability group as compared on total reading achievement scores measured by the Metropolitan Achievement Tests: Elementary Form F administered in October 1977.

Hypothesis II. There was a significant difference between an average reading ability group and a high reading ability group as compared on total reading achievement scores measured by the Metropolitan Achievement Tests: Elementary Form F administered in October 1977.

Hypothesis III. There was a significant difference between a low reading ability group and an average reading ability group as compared on science achievement scores measured by the Metropolitan Achievement Tests: Intermediate Form G administered in April 1980.

Hypothesis IV. There was a significant difference between an average reading ability group and a high reading ability group as compared on science achievement scores measured by the Metropolitan Achievement Tests: Intermediate Form G administered in April 1980.

Hypothesis V. There was no significant difference between total reading achievement gains and science achievement gains for a low reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

Hypothesis VI. There was no significant difference between total reading achievement gains and science achievement gains for an average reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

Hypothesis VII. There was no significant difference between total reading achievement gains and science achievement gains for a high

reading ability group as measured by the Metropolitan Achievement Tests: Intermediate Forms F and G administered in October 1978 and April 1980 respectively.

Hypothesis VIII. There was a significant difference between low reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

Hypothesis IX. There was no significant difference between average reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

Hypothesis X. There was no significant difference between high reading ability students with total reading achievement scores of 6.0 or above and those students scoring below 6.0 when compared on attainment of grade level science achievement in April 1980.

Hypothesis XI. There was no significant difference between low reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

Hypothesis XII. There was a significant difference between average reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

Hypothesis XIII. There was no significant difference between high reading ability students with total reading achievement scores of 7.0 or above and those students scoring below 7.0 when compared on attainment of grade level science achievement in April 1980.

Consideration of the action taken with regard to the thirteen hypotheses and the visual inspection of the data treated therein brings forward the following conclusions:

1. Science achievement as measured by the Metropolitan Achievement Tests was affected by total reading achievement.

2. Gains in science achievement scores and total reading achievement scores tended to progress concurrently.

3. Science achievement scores were adversely affected when students were required to respond to an achievement test with an estimated readability level that exceeded the total reading achievement level of the individual student.

4. Science achievement scores were adversely affected when students were required to study a basal elementary school science textbook which had an estimated readability level which exceeded the total reading achievement level of the individual student.

III. IMPLICATIONS

Classroom teachers should be cognizant of the relationship that exists between total reading achievement and science achievement as demonstrated by this study. Knowledge of this relationship is of particular importance to a school system which has adopted an elementary school science textbook series that is reading dependent.

The findings of this study indicated that science achievement tended to progress concurrently with total reading achievement. Therefore appropriate emphasis placed on remedial and/or developmental reading would be beneficial toward increasing science achievement scores of students who performed below grade level in science achievement.

A most important area of concern for the classroom teacher is the matching of science materials to the learners. The findings of this study demonstrated the importance of the student being expected to read both the achievement test used to measure science achievement and the basal elementary school science textbook series adopted by the school system, if grade level science achievement is to be attained.

The classroom teacher should consider the levels of total reading achievement demonstrated by the students to be tested before selecting an achievement test. The achievement test chosen should have an estimated readability level that is appropriate for the students, if they are to be expected to demonstrate their ability in science through science achievement scores.

The selection of a basal elementary school science textbook series should follow the same matching of materials to learners format as was outlined for the selection of an achievement test. Since textbook series are adopted for a five year period the possibility of a mismatch of materials to learners during the adoption period is likely. The use of vocabulary word lists, the rewriting of selected materials to a level nearer the reading abilities of the learners, or the selection of an alternate textbook written at a lower estimated readability level would be beneficial to the students in maintaining science achievement progress.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ahmann, Stanley J., et al. "Science Achievement: The Trend is Down," The Science Teacher, 42 (September 1975), 23-25.

Baldwin, R. Scott, and Rhonda K. Kaufman. "A Concurrent Validity Study of the Raygor Readability Estimate," Journal of Reading, 22 (November 1979), 148-153.

Belden, Bernard R., and Wayne D. Lee. "Readability of Biology Textbooks and the Reading Ability of Biology Students," School Science and Mathematics, 61 (December 1961), 689-693.

Belden, Bernard R., and Wayne D. Lee. "Textbook Readability and Reading Ability of Science Students," The Science Teacher, 29 (April 1962), 20-23.

Bennett, Lyle Gene. "An Analysis of the Readability of Sixth Grade Science Textbooks Using the Dale-Chall Formula and the Cloze Procedure Test," Dissertation Abstracts, 40 (1980), 5299A-5300A.

Bloom, Benjamin S. "Implications of the IEA Studies for Curriculum and Instruction," Educational Policy and International Assessment: Implications of the IEA Surveys of Achievement. Berkeley, California: McCutchan Publishing Corporation, 1975.

Blue, Larry Lamar. "A Study of the Influence of Certain Factors in Science Materials on the Reading Comprehension of Seventh Grade Pupils," Dissertation Abstracts, 25 (1965), 5625.

Bond, Guy L., and Miles A. Tinker. Reading Difficulties, Their Diagnosis and Correction. New York: Appleton-Century-Crofts, 1957.

Bormuth, John. "Comparable Cloze and Multiple Choice Comprehension Test Scores," Journal of Reading, 10 (February 1967), 291-299.

Brown, Walter R. "Science Textbook Selection and the Dale-Chall Formula," School Science and Mathematics, 65 (February 1965), 164-167.

Burkey, Jacob E. "The Readability of Flementary Science Materials," Dissertation Abstracts, 14 (1954), 1328.

Burmeister, Lou E. "A Chart for the New Spache Formula," The Reading Teacher, 29 (January 1976), 384-385.

Carin, Arthur A., and Robert B. Sund. Teaching Modern Science. Columbus, Ohio: Charles E. Merrill Publishing Company, 1970.

Carter, Glenda S., and Ronald D. Simpson. "Science and Reading: A Basic Duo," The Science Teacher, 45 (March 1978), 18-21.

Cramer, Ward, and Suzanne Dorsey. "Science Textbooks: How Readable Are They? Elementary School Journal, 70 (October 1969), 28-33.

Curtis, Francis D. Some Values Derived from Extensive Reading of General Science. New York: Teachers College Press, Columbia University, 1924.

Curtis, Francis D. "A Study of the Vocabulary of Scientific Articles Appearing in Daily Newspapers," School and Society, 23 (June 26, 1926), 821-824.

Curtis, Francis D. Investigations of Vocabulary in Textbooks of Science for Secondary Schools. Boston: Ginn and Company, 1938.

Curtis, Francis D. "Investigations of Vocabulary in Textbooks of Science for Secondary Schools," Research in Science Education 1938 through 1947, Robert W. Boenig, Editor. New York: Teachers College Press, Columbia University, 1979.

Dale, Edgar. Building a Learning Environment. Bloomington, Indiana: Phi Delta Kappa, Inc., 1972.

Dale, Edgar, and Jeanne L. Chall. "A Formula for Predicting Readability," Educational Research Bulletin, 27 (January 1948), 11-20, 28.

Dale, Edgar, and Jeanne L. Chall. "A Formula for Predicting Readability Instructions," Educational Research Bulletin, 27 (February 1948), 37-54.

Dale, Edgar, and Jeanne Chall. "The Concept of Readability," Elementary English, 26 (January 1949), 19-26.

Dale, Edgar and Ralph Tyler. "A Study of the Factors Influencing the Difficulty of Reading Materials for Adults of Limited Reading Ability," Library Quarterly, 4 (July 1934), 384-412.

Denslow, Orriene D. "Vocabulary and Sentence Study of Eight First Grade Science Books," Elementary English, 38 (November 1961), 487-490.

Dorst, Walter N. et al. Metropolitan Achievement Tests: Intermediate Form F. New York: Harcourt Brace Jovanovich, 1970.

Dorst, Walter N. et al. Metropolitan Achievement Tests: Intermediate Form G. New York: Harcourt Brace Jovanovich, 1970.

Dorst, Walter N. et al. Metropolitan Achievement Tests: Elementary Form F. New York: Harcourt Brace Jovanovich, 1972.

Driver, Jean Ricks. "The Intercorrelations among Three Readability Formulas When Applied to Selected Fourth, Fifth, and Sixth Grade Science Textbooks," Dissertation Abstracts, 32 (1972), 5999A.

Farr, James, James Jenkins, and Donald Paterson. "Simplifications of Flesch Reading Ease Formula," Journal of Applied Psychology, 35 (October 1951), 333-337.

Ferguson, Richard L. "ACT Scores—What's behind the Decline?" NASSP Bulletin, 60 (May 1976), 71-77.

Field, Maurice Houston et al. "The Approximate Reading Level Ranges and Means by Grade Level for Twelve Elementary School Science Textbook Series," Science Education Bulletin 101, Martin, Tennessee: Department of Elementary and Secondary Education, The University of Tennessee at Martin, 1975.

Flesch, Rudolf. Marks of Readable Style. New York: Bureau of Publications, Teachers College, Columbia University, 1940.

Flesch, Rudolf. "Estimating Comprehension of Difficulty of Magazine Articles," Journal of General Psychology, 27 (January 1943), 63-80.

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

Flesch, Rudolf F. The Art of Readable Writing. New York: Harper and Brothers Publishers, 1949.

Flesch, Rudolf F. How to Write, Speak, and Think More Effectively. New York: Harper and Brothers, 1960.

Flesch, Rudolf F. How to Test Readability. New York: Harper and Brothers, 1951.

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

Fry, Edward. Reading Instruction for Classroom and Clinic. New York: McGraw-Hill Book Company, 1972.

Fry, Edward. "A Readability Graph for Librarians, Part I," School Libraries, 19 (Fall 1969), 13-16.

Funk, James H. "What Research Says to the Teacher: An Examination of the Levels of Questions on Standardized Tests for Elementary Science," Science and Children, 14 (January 1977), 30-32.

Gilbert, Charles D. "An Examination of Readability Levels for Selected Basic Science Texts," School Science and Mathematics, 73, (December 1973), 747-758.

Good, Carter V. Dictionary of Education. Second Edition, New York: McGraw-Hill Book Company, 1959.

Gray, William, and Bernice Leary. What Makes a Book Readable? Chicago: University of Chicago Press, 1935.

Grove, Martha Judin. "Sample Frequency Modifications and Comparative Readability of Four Readability Formulas as Applied to Selected Textbooks on the Teaching of Reading," Dissertation Abstracts, 40 (1979), 1377A.

Gunning, Robert. The Technique of Clear Writing. New York: McGraw-Hill Book Company, 1952.

Harnishfeger, Annegret, and David E. Wiley. "Achievement Test Scores Drop, So What?" Educational Researcher, 5 (March 1976), 5-12.

Harnishfeger, Annegret, and David E. Wiley. "The Marrow of Achievement Test Score Declines," Educational Technology, 16 (June 1976), 5-14.

Harris, Albert J., and Milton D. Jacobson. "A Comparison of the Fog, Spache, and Harris-Jacobson Readability Formulas for Primary Grades," The Reading Teacher, 32 (May 1980), 920-924.

Harrington, Roma Lenore, and George Greisen Mallinson. "An Investigation of Two Methods of Measuring the Reading Difficulty of Materials for Elementary Science," Science Education, 42 (December, 1958), 385-390.

Johnson, George R. "An Objective Method of Determining Readability," Journal of Educational Research, 21 (April 1930), 283-287.

Kearney, Nolan D. "Sentence Length in 121 Representative First-Grade Readers," Journal of Educational Research, 38 (February 1945), 447-461.

Kennedy, Keith. "Reading Level Determination for Selected Texts," The Science Teacher, 41 (March 1974), 26-27.

Kirk, Roger E. Experimental Design: Procedures for the Behavioral Sciences. Belmont, California: Brooks/Cole Publishing Company, 1968.

Layton, James R. "A Chart for Computing the Dale-Chall Readability Formula above Fourth Grade Level," Journal of Reading, 23 (December 1980), 239-244.

Lee, Wayne D., and Madge E. Hislop. "Problems Involved in Placement of Students in a Biology Class," School Science and Mathematics, 68 (June 1968), 473-476.

Lewerenz, Alfred A. "Measurements of Difficulty of Reading Materials," Educational Research Bulletin, 8 (January 1929), 11-16.

Lively, Bertha, and S. L. Pressey. "A Method for Measuring the Vocabulary Burden of Textbooks," Educational Administration and Supervision, 9 (October 1923), 389-398.

Lorge, Irving. "Predicting Reading Difficulty of Selections for Children," Elementary English Review, 16 (October 1939), 229-233.

Lorge, Irving. "The Lorge and Flesch Readability Formulae: A Correction," School and Society, 67 (February 21, 1948), 141-142.

Mallinson, George G. "Reading in the Sciences: A Review of the Research," Reading for the Content Areas. New York: International Reading Association, 1972.

Mallinson, George G., Harold E. Sturm, and Lois M. Mallinson. "The Reading Difficulty of Textbooks in High School Biology," American Biology Teacher, 12 (November 1950), 151-156.

Mallinson, George Greisen, Harold E. Sturm, and Lois M. Mallinson. "The Reading Difficulty of Textbooks in Junior High School Science," The School Review, 58 (December, 1950), 536-540.

Mallinson, George G., Harold E. Sturm, and Lois M. Mallinson. "The Reading Difficulty of Textbooks for General Science," School Review, 40 (February 1952), 94-98.

Mallinson, George G., Harold E. Sturm, and Lois M. Mallinson. "The Reading Difficulty of Textbooks for High School Physics," Science Education, 36 (February 1952), 19-23.

Mallinson, George G., Harold E. Sturm, and Lois M. Mallinson. "The Reading Difficulty of Textbooks for High School Chemistry," Journal of Chemical Education, 29 (December 1952), 629-631.

Mallinson, George G., Harold E. Sturm, and Lois M. Mallinson. "The Reading Difficulty of Textbooks for General Physical Science and Earth Science," School Science and Mathematics, 54 (November 1954), 612-616.

Mallinson, George Greisen, Harold E. Sturm, and Lois Marion Mallinson. "The Reading Difficulty of Unit-Type Textbooks for Elementary Science," Science Education, 39 (December 1955), 406-410.

Mallinson, George Greisen, Harold E. Sturm, and Lois M. Mallinson. "The Reading Difficulty of Some Recent Textbooks for Science," School Science and Mathematics, 57 (May 1957), 364-366.

Mallinson, George Greisen, Harold E. Sturm, and Robert E. Patton. "The Reading Difficulty of Textbooks in Elementary Science," Elementary School Journal, 50 (April 1950), 460-463.

Mallinson, George G. et al. "The Reading Difficulty of Science Textbooks," Research in Science Education 1948 through 1952, J. Nathan Swift, Editor. New York: Teachers College Press, Columbia University, 1969.

Mills, R. E., and Jean R. Richardson. "What Do Publishers Mean by 'Grade Level'?" The Reading Teacher, 16 (March 1963), 359-362.

Moore, Arnold J. "The Preparation and Evaluation of Unit Text Materials in Science for Low Ability Junior High School Students," Science Education, 46 (December 1962), 488-492.

McLaughlin, G. H. "SMOG Grading: A New Readability Formula," Journal of Reading, 11 (May 1969), 639-646.

National Assessment of Educational Progress. Three National Assessments of Science: Changes in Achievement, 1969-77. Science Report No. 08-5-00. Denver: Education Commission of the States, 1978.

National Assessment of Educational Progress. National Assessments of Science, 1969 and 1973: A Capsule Description of Changes in Science Achievement. Science Report No. 04-5-00. Washington U.S. Government Printing Office, 1975.

Nelson, Donald J. "What's behind the Test Score Decline?" NASSP Bulletin, 60 (May 1976), 78-82.

Newport, John F. "The Readability of Science Textbooks for Elementary Schools," Elementary School Journal, 66 (October 1965), 40-43.

Ottley, LeRoy. "Readability of Science Textbooks for Grades Four, Five, Six," School Science and Mathematics, 65 (April 1965), 363-366.

Piltz, Albert, and Robert Sund. Creative Teaching of Science in the Elementary School. Boston: Allyn and Bacon, 1970.

Pooley, Robert C. "Reading and the Language Arts, Development in and through Reading," The Sixteenth Yearbook of the National Society for the Study of Education, Part I, 1961.

Rammell, Joseph A. "An Objective Study for the Basis of Selecting an Eighth Grade General Science Textbook," Science Education, 47 (April 1963), 258-264.

Ramsey, George A., and Robert W. Howe. "An Analysis of Research Related to Industrial Procedures in Elementary School Science," Science and Children, 6 (April 1969), 25-36.

Raygor, Alton L. "The Raygor Readability Estimate: A Quick and Easy Way to Determine Difficulty," Reading: Theory, Research and Practice. The Twenty-sixth Yearbook of the National Reading Conference, 1977.

Ricker, Kenneth S. "But Can They Read? A Look at Readability Formulas," The Science Teacher, 45 (March 1978), 22-24.

Sapone, Carmelo V., and Joseph R. Giuliano. "The Test Score Decline: Are the Public Schools the Scapegoat?" Educational Technology, 16 (June 1976), 43-44.

Sax, Gilbert. Principles of Educational Measurement and Evaluation. Belmont, California: Wadsworth Publishing Company, 1974.

Sharp, Vicki F. Statistics for the Social Sciences. Boston: Little, Brown and Company, 1979.

Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, 1956.

Skymansky, James A., and Larry D. Yore. "Assessing and Using Readability of Elementary Science Texts," School Science and Mathematics, 79 (December 1979), 670-676.

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

Spache, George D. "The Spache Readability Formula," Good Reading for Poor Readers, Champaign, Illinois: Garrard Publishing Company, 1974.

Stauffer, Russell G. Directing the Reading-Thinking Process. New York: Harper and Row, 1969.

Stauffer, Russell G. "A Vocabulary Study Comparing Reading, Arithmetic, Health and Science Texts," The Reading Teacher, 20 (November 1966), 141-147.

Stephens, Robert Gerald. "Use of the Cloze Procedure as a Criterion for Evaluating the Applicability of Selected Readability Formulas to Science Reading Materials," Dissertation Abstracts, 32 (1972), 6020A.

Stevens, Kathleen C. "Readability Formulae and McCall-Crabbs Standard Test Lessons in Reading," Journal of Reading, 23 (January 1980), 413-414.

Taylor, "Cloze Procedure: A New Tool for Measuring Readability," Journalism Quarterly, 30 (Fall, 1953), 415-433.

Tennessee Department of Education. Tennessee Assessment of Educational Progress Results Report: Reading, 1974-75. Nashville, Tennessee: State Department of Education, 1975.

Tennessee Textbook Commission. Tennessee Official List of Textbooks with Wholesale Prices, Retail Prices and Copyright Dates. Nashville: Tennessee Book Company, 1979.

Thelen, Judith N. "Using the Cloze Test with Science Textbooks," Science and Children, 12 (November-December 1974), 26-27.

Thorndike, Edward L. A Teacher's Word Book. New York: Teachers College Bureau of Publications, Columbia University, 1921.

Thorndike, Edward L. A Teacher's Word Book of 20,000 Words. New York: Teachers College Bureau of Publications, Columbia University, 1931.

Thorndike, Edward L., and Irving Lorge. The Teacher's Word Book of 30,000 Words. New York: Teachers College, Bureau of Publications, Columbia University, 1944.

Vogel, Mabel, and Carleton Washburne. "An Objective Method of Determining Grade Placement of Children's Reading Material," Elementary School Journal, 27 (January 1928), 373-381.

Walker, Helen M., and Joseph Lev. Elementary Statistical Methods. New York: Holt, Rinehart and Winston, Inc., 1969.

Williams, David L. "The Effect of Rewritten Science Textbook Materials on the Reading Ability of Sixth-Grade Pupils." Unpublished doctoral dissertation, University of Illinois, 1964.

Williams, David L. "Rewritten Science Materials and Reading Comprehension," Journal of Educational Research, 61 (January 1968), 204-206.

Woffort, Barbara A. "Using Reading Miscue Analysis to Investigate Publishers' Suggested Readability Levels for Elementary Science Textbooks: A Comparative Study." Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, 1975.

Yoakum, Gerald A. Basal Reading Instruction. New York: McGraw-Hill Book Company, 1955.

APPENDIX

DATA COLLECTION FORM

_____
Name

_____
No.

_____
No.

Year	Word Know.	Word Anal.	TL Read	Science	

VITA

Maurice Houston Field was born in Henderson, Kentucky, on April 16, 1938. He attended Atkins-Porter Elementary School and was graduated from E. W. Grove High School in June 1956. Both schools are located in Paris, Tennessee. In September 1956 he enrolled at The University of Tennessee at Martin, where he received the Bachelor of Science degree with a major in Secondary Education and an endorsement in Natural Science in August 1959. The following September he enrolled at The University of Tennessee, Knoxville, and received a Master of Science degree with a major in Science Education in August 1960.

Between September 1960 and August 1962 he taught chemistry and biology at White Station High School in Memphis, Tennessee. He also coached basketball, football, baseball, and golf. The golf team won the state tournament in 1961. During this same period he pursued the Master of Arts degree in Educational Administration and Supervision at Memphis State University. The degree was completed in August 1962.

The author has been employed at The University of Tennessee at Martin since 1964, where he is an Associate Professor and Chairman of the Department of Elementary Education and Related Services. He is a member of Phi Kappa Phi, Phi Delta Kappa, Science Association of Tennessee and is a life member of the National Science Teachers Association. He has served as President of the Science Association of Tennessee and has twice been the program chairman for the annual meeting.

He is the son of Maurice Elmer Field and the late Louise Merrick Field of Paris, Tennessee. He is married to the former Carla Jane Wilson of Memphis and has two children, Alicia Elaine and Maurice Edward.