A NEW METHOD FOR DETERMINING THE READABILITY

OF STANDARDIZED TESTS USED IN COUNSELING

by

Fritz W. Forbes
A.B., University of Kansas, 1936
M.A., University of Kansas, 1946

Submitted to the Department of Education
and the Faculty of the Graduate School
of the University of Kansas in partial
fulfillment of the requirements for the
degree of Doctor of Education.

Advisory Committee:

Redacted Signature
_____

Redacted Signature
_____

Redacted Signature
_____

July, 1952

ACKNOWLEDGMENTS

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF TABLES

CHAPTER I

INTRODUCTION

Importance of the Study

Since the early nineteen-hundreds, more and more
emphasis is being placed upon objective measurement of
readability. Early attempts to evaluate the reading diffi-
culty of textbooks have brought about this recent concern
about readability. Interest in objective measurement of
readability has been growing steadily since the first
basic research was done by Vogel and Washburne.[1] It is
estimated by Klare[2] that up to the present time there are
thirty-four formulas or methods for determining the reading
difficulty of printed material. Five of the more recently
developed formulas have been singled out for critical
analyses here.

Scientific interest in reading is becoming
important to teachers of all levels. The elementary
teachers have for many years recognized the value of
standardized instruments for measuring reading ability and

---

1 Mable Vogel and Carleton Washburne. "An Objective Method
   of Determining the Grade Placement of Children's Reading
   Material." The Elementary School Journal, 28 (January
   1928) 373-381.

2 George R. Klare. "Evaluation of Quantitative Indices
   of Comprehensibility in Written Communication." Un-
   published Ph.D. Thesis, 1950, University of Minnesota.

for grading reading books. Secondary teachers are becoming aware that they too must meet the reading needs of their pupils.

New and varied problems have come with the change of the school population. Many children are now in school who a few years ago would not have been expected to attend. Many more retarded pupils have come with the influx of these boys and girls. They make it necessary that special provisions be made for remedial instruction in reading. At the same time, textbook materials have to be rewritten to fit the lower levels of pupil reading ability. If the schools are going to hold these pupils beyond compulsory educational age, they must adapt their programs to the pupil needs and educate each according to his abilities and potentials.

This varied and increased public school population has created a need to know just where these pupil abilities lie, what can be expected of such pupils and where they will fit best into society. This is the field of student personnel that has grown amazingly within the past decade. It meant that tests had to be developed to identify the pupil and try to point out his needs. The standardized tests, check lists and inventories used in any of the five major areas of such testing must be understandable to the person taking them if the scores are to be valid. These areas of testing, personality, achievement, interest,

intelligence, and aptitude must present tests that are readable to the people taking them.

Little has been done to ascertain the reading level necessary to understand the content of these standardized testing materials. Johnson and Bond[1] have one of the few articles in this specific area. In their paper a single formula was used for testing reading ease of nine standardized tests. The general conclusion was that tests are being administered to people who do not understand them because the reading level of the tests is too high.

Stefflre[2] made a study of the relative reading difficulty of six interest inventories using the Flesch formula. High correlation between the Flesch formula and other formulas was claimed. Roeber[3] compared seven interest inventories as to word usage. The percentage of occurrence of different words appearing in the inventories was computed. He found a large number of words beyond the understanding of ninth graders. Thus, his recommendation for a glossary of terms does appear in a later form of one of the inventories.

1 Ralph H. Johnson and Guy L. Bond. "Reading Ease of Commonly Used Tests." Journal of Applied Psychology, 34 (October 1950) 319-324.

2 B. Stefflre. "The Reading Difficulty of Interest Inventories." Occupations, 26 (November 1947) 95-96.

3 Edward C. Roeber. "A Comparison of Seven Interest Inventories with Respect to Word Usage." Journal of Educational Research, 42 (September 1948) 8-17.

Testing instruments are becoming so varied and numerous that users of them need every help possible to determine the usefulness of the instruments. One aid to counselors is the objective measurement of readability--of which little is available at this time. Advocates of formulas or methods for determining readability thus have made no attempt to determine the reading difficulty of standardized testing material. Formulas have been based principally upon recreational reading.

The various formulas have been devised to estimate the difficulty of a specific type of reading matter. When the formulas are applied to their respective reading material they each give classification into broad relative order. When the formulas are applied carefully for estimating a single aspect of difficulty, they are used correctly. A device such as a readability formula lends itself to misure. Formulas designed for determining the reading difficulty of certain types of printed matter are used for estimating the difficulty of all reading material. These formulas are no "cure all." At the best they are yard sticks, and when correctly applied to their specific field will give a relative order classification of the material. The use of such devices is not a substitute for common sense and experience.

Methods of Attack

Many and varied factors have been used in attempts to determine the readability of printed material. It is beyond the scope of this study to give a critical analysis of all the various factors previously used to determine readability. Detailed analysis will be made later of the factors employed by the five formulas used in this study.

The more common variables used to predict readability are:

1. Number of running words.

2. Percentage of infrequent, uncommon, or hard words.

3. Vocabulary diversity.

4. Some weighted measure of vocabulary difficulty.

5. Average sentence length.

6. Percentage of polysyllabic words.

7. Number of abstract words.

8. Number of affixed morphemes (prefixes, suffixes, etc.)

9. Percentage of prepositional phrases.

10. Percentage of indeterminate clauses.

11. Number of simple sentences.

12. Number of personal pronouns.

13. Number of words expressing human interest.

14. Percentage of colorful words.

15. Number of words representing fundamental life experiences.

16. Percentage of words beginning with certain letters.

The first five of the above list are used most often. All five are used singly or in combination by the formulas studied in this paper. Other factors will be enumerated in the development of the review of the literature.

Materials Tested

A number of practical applications have been made of the various techniques for judging readability. It will suffice here to give only a brief mention of the author of the techniques and the work done by him. Allard[1] determined the difficulty of poems which are commonly presented to grade school pupils. Yoakam[2] in working out his readability formula ascertained the grade placement of a number of school readers published between 1930-39, later thirty-three school readers published between 1940-45, and twenty well-known children's books were rated according to his formula. The Department of Agriculture, Extension Service

---

1 J. A. Allard. "Difficulty of Poems Commonly Presented to Elementary School Pupils." University of Pittsburgh Bulletin, 42 (October 1945) 9-18.

2 Gerald A. Yoakam. "The Reading Difficulty of School Textbooks." The Elementary English Review, 22 (December 1945) 304-309.

Division,[1] analyzed their publications as to reading difficulty, taking into account the reading ability of prospective readers. The Curriculum Division of Ohio State University under the direction of Chall[2] analyzed books on philosophy to determine the grade level. Brittanica Junior Encyclopedia, Compton's Encyclopedia, and the World Book were given a reading difficulty rating determined by Edgerton.[3] Berger[4] made a reading difficulty analysis of nine third grade health readers. Later, Dale and Chall[5] devised their formula on adult health reading material. The John Newberry prize books were given reading grade placements determined by Miller.[6] Religious tests have been critically analyzed by Latimer[7] in an unpublished thesis from Pittsburgh

1   Amy Bronna Cowing. "They Speak His Language." Journal of Home Economics, 37 (October 1945) 478-489.

2   Jeanne S. Chall. "This Business of Readability." The Elementary School Journal, 47 (January 1947) 492-496.

3   Ronald Edgerton. "How Difficult are Children's Encyclopedias? I and II." The Elementary School Journal, 45 (March 1945) 378-385.

4   Herman I. Berger. "The Difficulty of Third-Grade Health Readers." The Elementary School Journal, 47 (March 1947) 391-395.

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

6   Leo R. Miller. "Reading-Grade Placement of the First Twenty-Three Books Awarded the John Newberry Prize." The Elementary School Journal, 46 (March 1946) 394-399.

7   Edward H. Latimer. "A Comparative Study of Recent Techniques for Judging Readability." Unpublished Ph.D. Thesis, University of Pittsburgh, 1947.

University. Brayfield and Reed[1] determined the readability
of occupational information booklets. Most recently Union-
Management Agreements have been studied by Tiffin and Walsh.[2]
About the same time Crissy[3] attempted to answer the question:
"How readable are employee handbooks?"

No attempt has been made thus far to develop a
readability technique for use exclusively with standardized
testing materials. Critical analysis is being made in the
present study of the five more recent formulas in order to
devise a method suited for evaluating the reading difficulty
of such materials.

Definition of Terms

Readability

The word readability has become common to the
English language, but what do we mean by the term read-
ability or a readable book? Webster's Unabridged Dictionary
defines readable as: "legible," "easy to read because of
interesting or pleasing," "that which permits or admits of

---

1 Arthur H. Brayfield and Patricia A. Reed. "How Readable
are Occupational Information Booklets?" Journal of
Applied Psychology, 34 (October 1950) 325-328.

2 Joseph Tiffin and Francis X. Walsh. "Readability of
Union-Management Agreements." Personnel Psychology, 4
(Winter 1951) 327-337.

3 William J. E. Crissy. "How Readable are Employee Hand-
books?" Personnel Psychology, 4 (Winter 1951) 383-395.

reading." Much confusion could result if we stopped here.

Gray and Leary[1] asked a group of librarians, publishers, and teachers what they thought makes a book readable. Hundreds of descriptive statements were received agreeing that content was the most important item. Factors of style, format, and organization followed in the order named.

Leary[2] a few years later summarized the survey:

> ...According to the combined opinion of these judges, then, if you give a reader a theme that interests him, whether it concerns people, travel, adventure, science, or business, you have made a strong attack upon the problem of readability. If in addition, you discover what style of expression is best suited to the reader's needs and tastes, that is, the scope of vocabulary and the kind of sentences which he reads easily, and the type of approach that pleases him, you have the final solution of the problem close at hand. In the opinion of these judges the attractiveness of the book, its mechanical set-up, and its general plan of organization are matters of minor importance.

The typographical aspects of readability have been investigated by Tinker and Paterson.[3] In their book,

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

2 Bernice E. Leary. "Difficulties in Reading Material." Reading in General Education. American Council on Education, 1940, p. 280.

3 Miles A. Tinker and Donald G. Paterson. "Reader Preference and Typography." Journal of Applied Psychology, 26 (January 1942) 38-40.

What Makes Type Readable, their general conclusion was:
"In general, we have used the words legibility and
readability interchangeably to mean ease and speed of
reading printed material at a natural distance." The
size of the print did not make as noticeable difference
in reading ease as did the spacing of the letters.
Burtt[1] also found that it took greater effort to read
capital letters than lower case letters.

Interest also plays an important role in
readability. Surprise, liveliness, and animalness seem
to hold the attention of young children, according to
Gates.[2] Zeller[3] found that action and humor exerted
greater influence on junior high students. Sterner[4] found
that among high school students "adverture is the favorite
with adolescents, humor is a close second, and the love
theme is very popular with high school girls." Adult

---

1 Harold E. Burtt. "Typography and Readability." Read-
ability, Reprint from Elementary English, (January-May
1949) 26-35.

2 Arthur I. Gates. Interest and Ability in Reading.
Chicago: The Macmillan Company, 1930. p. 89-90.

3 Dale Zeller. Relative Importance of Factors of Interest
in Reading Materials of Junior High Pupils. Contributions
to Education, No. 841. New York: Bureau of Publications,
Teachers College, Columbia University, 1941. p. 73.

4 Alice P. Sterner. Radio, Motion Pictures and Reading
Interests. Contributions to Education, No. 932. New
York: Bureau of Publications, Teachers College, Columbia
University, 1947.

interests vary greatly according to their socio-economic status.

Another aspect of readability is comprehension--that is, printed material is readable by any certain group when it can be understood by this group, when the words give a meaning. The understanding of printed material has been tested only by asking questions about the content read. Comprehension, then, is the ability to answer specific questions. Obviously, if we select reading material for the majority it will be more difficult than if we had chosen reading comprehensible to all.

There is no composite measure of all phases of readability. We can only consider separately the aspects of typography, interest, and comprehension; then make a judgement, based upon common sense and experience, as to a particular reading piece for a specific group of readers.

Objective Evaluation

The purpose of this research was to devise a measure of reading difficulty sufficiently refined and easily applicable so as to make possible identical results by two or more evaluators working independently on the same standardized testing instrument. This technique was in contrast to previous methods which relied to a large extent upon personal opinion or editorial judgment.

Vocabulary

The understanding of words in reading is basic, for without the knowledge of the meaning of words there can be no reading. Reading is the comprehension of meaning from the printed page through seeing relationships between words. Good readers are not conscious of single words; their training and practice allow them to get meanings from phrases, sentences, and paragraphs.

Most vocabulary lists are assumed to be "meaning" lists. This is not necessarily true. Many words are recognized by sight without clear understanding as to their full meaning, especially words having multiple meanings. We would expect such words as run and set to be recognizable by most advanced readers; yet, consider that the American College Dictionary published by Random House in 1947 gives 104 numbered definitions for run and 68 for set.

In order to determine if there is a relationship between "the readability of pupils' composition and their measured intelligence," Lorge and Kruglov[1] secured data from fifty high school pupils. An analysis of the records showed that "the structural elements of written

---

1 Irving Lorge and Lorraine Kruglov. "The Relationship Between the Readability of Pupils's Compositions and Their measured Intelligence." Journal of Educational Research, 43 (February 1950) 467-474.

expression are not related significantly to measured intelligence" when educational level is held constant. It was found, also that high school pupils write "at a level two grades below their understanding level," thus supporting the hypothesis that there is a difference "between the expressional level and the level of understanding for the same person."

In this study no attempt was made to distinguish between meanings of words with multiple definitions nor was any attempt made to place the words in a grade placement scale. Vocabulary was also taken to mean words as individual units, not a series of symbols producing complete sentences.

Statement of the Problem

Concurrent with the development of numerous standardized testing instruments, the expanding use of standardized tests in counseling, and the concern about readability is the demand for more effective means of determining the usability of such tests.

It was the purpose of this study to develop an objective measure of readability of standardized tests commonly used in counseling, thus allowing the counselor to more effectively determine the usefulness of such instruments.

Several formulas have been devised based upon various factors of readability. None of these have been evolved specifically for measuring the readability of standardized tests. However, each uses the factor of word difficulty to some extent in its measure of readability.

By applying five of the more recent formulas to the commonly used standardized testing instruments, a mean score of difficulty was obtained.

The reading difficulty of these tests is one primarily of vocabulary, since there is no continuity of thought or theme carrying through the reading material. A vocabulary burden for each test was figured and ranked in order of difficulty to approximate as nearly as possible the mean score obtained from having applied the five formulas to the selected tests. The vocabulary burden was figured from weights assigned to each difficult word; this weight being taken from some accepted word list. This weighted score gave the grade level of difficulty of the vocabulary used in standardized tests.

Summary

Objective measurement of reading difficulty of standardized tests is a neglected area of readability. The thirty-four formulas and techniques that have been developed up to the present time have failed to consider the

readability of these testing instruments. The purpose of
this paper is to develop a method specifically for
determining the reading difficulty of standardized tests
used commonly for counseling purposes; thus making it
possible for the users of these instruments to select
those tests which can be read and understood by the people
taking them.

CHAPTER II

A SURVEY OF THE LITERATURE RELATED TO THE PROBLEM

Introduction

The review of the literature has been divided into five sections: Annual Summaries and Bibliographies, Readability, Vocabulary, Reading Level, and Studies Related Specifically to the Study. Each of these areas of the literature of reading will be reviewed separately.

Annual Summaries and Bibliographies

The history of the scientific study of readability has been carefully outlined in recent studies. William S. Gray, University of Chicago, is a most prolific worker in the general field of reading research. Annual Summaries of reading Investigations appear each year in the _Journal of Educational Research._ Prior to these annual summaries a volume entitled, _Summary of Investigations Relating to Reading_[1] appeared which included 436 studies published prior to July 1, 1924. During the subsequent twenty-seven years a total of 2,548 studies have been reported by Gray in his annual summaries. These annotated bibliographies are divided into three headings: Sociology of Reading,

---

1 William S. Gray. _Summary of Investigations Relating to Reading._ University of Chicago Monographs, No. 28. Chicago: University of Chicago Press, 1925.

Physiology and Psychology of Reading, and the Teaching of Reading.

Under the title of "Frontiers in Educational Research in Reading," Gates[1] discussed certain findings as revealed by research and defined nine major needs for further research:

1. "Research on the characteristics and components of reading." More needs to be known of the relation between reading and reasoning.

2. "Comparison of reading and other media of learning." What is the value of diagrams, sound pictures, and mechanical contrivances to effective reading?

3. "Research on opportunities and needs present by society in the near future." Because of the diversity of languages, more pictorial and non-verbal communications for international understanding is needed.

4. "Research on the value of reading and other means of learning in school programs." The programs in school have been developed primarily from printed reading materials; while television, radio, and pictures have developed outside of school.

5. "Research on the improvement of the organization and character of reading materials for different purposes." Schools and texts should take a lesson from tabloids and comics.

6. "Research on individual ability and limitations in learning by reading and in other ways."

7. "The need for experimentation to counteract the effect of specialization in research in reading." During the last two decades the amount of research on reading has extended beyond that in any other media of learning. Research in reading has become a very specialized field. "Although many reading specialists are familiar with curriculum, theory, and school practices, there are many others who are quite innocent of knowledge of modern methods and purposes of the school."

8. "Need for critical review by and cooperation with other specialists." This cooperative review could be with clinical psychologists, physicians, and social workers.

9. "Need of more studies of general theories of learning, etc., implicit in results of research on reading.

A broad outline of the progress made in reading research between the years 1930-40 was published by Traxler[1] in 1941. These ten years of research revealed that great strides were made during this time. During the nineteen-thirties, laboratory investigation continued to make important contributions to our knowledge of reading. Important studies came from the Psycho-Educational Clinic

---

1 Arthur E. Traxler. Ten Years of Research in Reading. Educational Records Bulletin, No. 32, Educational Records Bureau, 1941.

under Dearborn's direction at Yale, Buswell's laboratory
at the University of Chicago, Gates and his associates
at Columbia, Tinker and others at the University of
Minnesota, Eames at Boston University, the Dartmouth Eye
Institute of the Dartmouth Medical School, Betts' Clinic
at the Oswego Normal School, and later at the Pennsylvania
State College.

Traxler and Townsend[1] made a most comprehensive
survey of research done in the field of reading between
1940-50. The major part of this volume was devoted to an
annotated bibliography of 527 items. Essential findings
were summarized under twenty headings; most of the areas
of investigation were continuations of work begun earlier
and extended during the period covered by these five years
of research.

These fifteen years of investigations into
reading research, combined with more recent studies, were
included in a single paper for presentation at an Institute
of Methods of Teaching Reading in Germany by Traxler.[2]
This paper on research in reading reviewed briefly the

---

1 Arthur E. Traxler and Agatha Townsend. Another Five
   Years of Research in Reading. Educational Records
   Bulletin, No. 46, Educational Records Bureau, 1946.

2 Arthur E. Traxler. "Research in Reading in the United
   States." Journal of Educational Research, 42 (March
   1949) 481-499.

nature and scope of early studies in reading, and the broad
outlines of reading progress between 1930-48.

Readability

In addition to the annual summaries published by
Gray in the _Journal of Educational Research_ covering
several phases of the reading problem, a more recent
article appeared summarizing the progress made in the study
of readability.[1] In this investigation he listed the
studies of readability under two headings: vocabulary
studies and the use of various factors in determining read-
ing difficulty. Under the first title, he included the
methods of Lively and Pressey (1928), Vogel and Washburne
(1928), Lewerenz (1929), Johnson (1930), Patty and Painter
(1931), and Yoakam (1938). In the second group, Gray
listed the studies by Vogel and Washburne (1928), Dale
and Tyler (1934), Gray and Leary (1935), Lorge (1938),
and Flesch (1943).

Chall[2] outlined the history of attempts to measure
readability. Credit was given here to the following names
for pioneering in the study of readability: E. L. Thorndike,

1 William S. Gray. "Progress in the Study of Readability."
   The Elementary School Journal, 47 (May 1947) 491-499.

2 Jeanne S. Chall. "This Business of Readability."
   Educational Research Bulletin, 26 (January 1947) 2-7.

Robinson, Lively and Pressey, Keboch, Dolch, Morphett (Vogel) and Washburne, McClusky, Lorge, Dale and Tyler, Gray and Leary, Morris and Holverson, and Flesch. Other names were listed in a bibliography that Chall prepared for interested readers giving the entire developmental history of the study of readability.

Four questions were discussed and answers sought through a critical review by Betts:[1]

1. What approaches have been made to the investigation of factors in readability? The present trend seems to be centered on language and the content of reading material. Workers have concentrated upon relationships between these factors: vocabulary difficulty, vocabulary diversity, sentence length and structure, "human interest," and meaning. By using combinations of certain of these factors, formulas have been derived for predicting difficulty of reading material. "Objective measures of readability are given precedence over author and teacher judgment."

2. "What factors contribute to current interest in readability?" Problems regarding readability have been brought to the front for a number of reasons: (1) The trend to emphasize reading as the chief aid to learning

1 Emmett A. Betts. "Readability: Its Application to the Elementary School." Journal of Educational Research, 42 (February 1949) 438-459.

appears to be on the increase, (2) A better professional understanding of the relationship between the readability of instructional materials and frustrations in reading situations is being sought, (3) discrepancies between grade scores achieved on standardized tests and the ability of students to read instructional material have been noticed by classroom teachers, (4) significant differences in reading difficulty of basal textbooks for a given grade level have been noted by both classroom teacher and research worker, (5) the trend to reduce the vocabulary load of basal textbooks has increased interest in readability, (6) the slow extension of practices in the direction of experience approach of learning has directed attention to readability, and (7) last, but not least, of the reasons for the current interest in readability is the problem of commercial value. Goods are sold via the speaking voice or printed page to the degree by which the commercials have been prepared in terms of the hearing and reading comprehension levels of the buying public.

3. "What have investigations yielded?" Twelve investigations on the prediction of readability were reviewed. Betts' findings and conclusions were that readability is significantly influenced by: average sentence length in number of words, the number of simple sentences, the number of prepositional phrases, percentage of different words in a selection, the number of uncommon

words in terms of word lists, number of words beginning
with certain letters, number of words with two or more
syllables, and the number of adjectives, adverbs, personal
pronouns, and other words related to human relationships.
Some structural elements tend to operate differently at
low levels than they do at high levels of readability. For
example, "uncommonness" of words appears to be a greater
handicap to children than to adults. Sentence length or
structure operates at both child and adult levels as a
factor in readability. In general, easy material has short
sentences and high proportion of common words, mono-
syllabic words, and personal references; difficult material
has long and complex sentences, and a high proportion of
uncommon words, different words, polysyllabic words, and
prepositional phrases.

4. "What uses can be made of these findings in
elementary schools?" The efforts of the overworked and
underpaid classroom teacher would be facilitated by read-
ability indices to books and to units and tests within the
book. Some agency, perhaps a National Bureau of Readability
Standards, should make itpossible to provide comparable
ratings on all instructional materials sold for use in
schools. This type of service would do two things: stimu-
late more research on this problem, and give teachers
dependable information on the relative difficulty of
instructional material.

A brief review of the investigations concerned with the prediction of readability that have come to the writer's attention is summarized here. The five formulas used in this research were fitted into their chronological order. More detailed explanation will be given later.

Lively and Pressey[1] reported one of the early significant studies in 1923. The attempt was made to establish the vocabulary burden of textbooks. Each word in a sample of one thousand word samples was assigned an index number from Thorndike's Teachers Word Book of 10,000 words. The high numbers in this word list indicated ease of reading. However, they stated that their findings gave only a critical index of frequency of words and not a measure of difficulty.

Keboch[2] in 1927 studied the word difficulty and variability of five history textbooks. A statistic based on the second five thousand words found in the Thorndike list was used.

The first regression equation for determining the readability of children's books was published by Vogel and

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

2 F. D. Keboch. "Variability of Word Difficulty in Five American History Textbooks." Journal of Educational Research, 40 (January 1927) 22-26.

Washburne[1] in 1928. The factors involved were: number of different words, number of uncommon words, number of prepositions, and the number of simple sentences. Uncommon words were taken to be all words not reported in Thorndike's Teachers Word Book. This so-called "Winnetka Formula" was tested against school grades three to eight by having the children report their likes and dislikes of the books they had read from the library. A correlation of .845 was found.

Lewerenz[2] in 1929 reported finding that words beginning with the letters b, h, and w tended to be "easy" and that those words beginning with the letters e and i were more difficult.

Johnson[3] in 1930 based his estimate upon a poly-syllabic word count. He found that the count of technical words, based on the Pressey Technical Vocabulary Lists, gave the same relative rating as did the count of poly-syllabic words. The comparative counts were made on four history books and three home economics textbooks.

1 Mabel Vogel and Carleton Washburne. "An Objective Method of Determining Grade Placement of Children's Reading Material." Elementary School Journal, 28 (January 1928) 373-381.

2 Alfred S. Lewerenz. "Measurement of the Difficulty of Reading Materials." Educational Research Bulletin (Los Angeles City Schools), 8 (March 1929) 11-16.

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

A technique for obtaining an index to difficulty, based on a weighted vocabulary sampling, was reported in 1931 by Patty and Painter.[1] All state adopted textbooks in Indiana were measured. The index of difficulty was obtained by dividing the weighted values of the sampled words by the number of different words in the sample. The weights for these values were the index numbers given by Thorndike in his Teachers Word Book. The sampling was done by taking the third line from every fifth page. The index of difficulty found was the ratio between the range of different words and the frequency of usage.

A composite of the number of different words, the Thorndike rating, and the number of polysyllables was used by Brown[2] to give a vocabulary rating of difficulty. This study was done with high school textbooks in 1931.

Holland[3] counted only the simple sentences in silent reading material to arrive at a rating for difficulty of reading. He found that the effect varied with different

1 W. W. Patty and W. I. Painter. "Improving Our Methods of Selecting High School Textbooks." Journal of Educational Research, 24 (June 1931) 23-32.

2 Robert Brown. "Vocabularies of History and Reading Textbooks." The Principal and Supervision, Tenth Year-book of the Department of Elementary School Principals. 1931, p. 408-411.

3 Benjamin F. Holland. "The Effect of the Length and Structure of Sentences on the Silent Reading Process." Psychological Bulletin, 30 (November 1933) 668-669.

patterns of sentences, with different individuals, and with different groups of subjects.

Lewerenz's doctoral dissertation[1] was completed in 1932 but his vocabulary grade placement formula did not receive wide circulation until 1935.[2] The approach by Lewerenz was based on three factors: vocabulary diversity, vocabulary difficulty, and vocabulary interest, each of which is evaluative in isolation to the others. Vocabulary diversity was obtained from a one thousand word sampling by identifying the different common words with a pre-established list of five hundred most frequently used words. Vocabulary difficulty was determined by calculating the percentage of polysyllabic words beginning with the letters b, h, w, e, and i in a thousand word sample. The vocabulary interest was established by counting the colorful words; these were compared with the colorful adverbs and adjectives in the Thorndike Teachers Word Book.

McClusky[3] found significant differences in comprehension related to sentence length, frequency of technical

---

1 Alfred S. Lewerenz. Techniques for the Objective Evaluation of the Vocabulary Used in Printed Matter. Unpublished Ph.D. Thesis, University of Southern California, 1932.

2 Alfred S. Lewerenz. "A Vocabulary Grade Placement Formula." Journal of Experimental Education, 3 (March 1935) 236.

3 Howard Y. McClusky. "A Quantitative Analysis of the Difficulty of Reading Materials." Journal of Educational Research, 28 (December 1934) 276-282.

terms, frequency of polysyllables, and especially to the
number of common concrete nouns. Several types of
material were used in this study of 1933; fiction,
political science, economics, sociology, psychology, and
physics. Units of one hundred words were selected and the
number of ideas in these units computed. The results
seemed to indicate, stated McClusky, that there was very
little difference in the number of ideas per hundred words
for the various passages.

In 1934 Dale and Tyler[1] carried on reading experi-
ments with colored people of limited reading ability. A
formula was based upon the number of certain previously
marked technical words, the number of non-technical words
and the number of indeterminate clauses. A correlation
between the score from this equation and reading compre-
hension test scores was .511. The non-technical word
element was defined by means of a list containing the 769
words which were common to the first thousand words of both
the Thorndike list and the International Kindergarten Union
list. The latter list registers the vocabulary of pre-
school children.

The Dale-Thorndike combined word list was used by

---

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

Ojemann[1] in 1934 for measuring vocabulary difficulty of adult education material. He used as the criterion a score indicative of the average reading ability of his subjects. The number of prepositional phrases and the length of sentences were found to give significant differences in reading difficulty, even when used without the vocabulary factor.

Gray and Leary[2] had in mind two questions to be answered when they published their monumental research: (1) What makes a book readable? and (2) Readable for Whom? An adult population of limited reading ability was selected of 1,690 adults from various geographical localities, occupational work, grade placement and age for the sampling. Five factors were finally selected from a list of eighty-two for a regression equation. They were: (1) number of different hard words (not included in the Dale list of 769 easy words), (2) number of first, second, and third person pronouns, (3) average sentence length in words, (4) percentage of different words, and (5) number of prepositional phrases. Predictions were made on approximately one hundred word samplings from each chapter

1   Ralph H. Ojemann. "The Reading Ability of Parents and Factors Associated with Reading Difficulty of Parent-Education Materials." Research in Parent-Education II. University of Iowa Studies in Child Welfare, 7 (1934) 9-32.

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

of a book. The specific conclusions arrived at from this study were: (1) adults in the upper one-sixth of the total sample could read more difficult types of material, (2) approximately one-half of the adults tested could read with reasonable ease and understanding general reading materials, (3) the lowest one-third of the 1,690 adults tested were unable to engage intelligently in adult reading activities, and (4) approximately one twenty-fifth could read only material of the second grade level.

The Morriss-Holverson Idea Analysis Technique[1] was completed in 1938. Words were grouped into four classifications: (1) classification I, simplest word labels representing fundamental or elemental experiences in the life of a people in a given culture, such as "mother," "water," "home," etc., (2) classification II, words also learned early in life which differ from classification I in being word-ideas which are localisms, such as "corn," "plow," "cattle," etc., (3) classification III, words signifying concrete ideas, such as "filament," "Van Gogh," "Iraq," etc., and (4) classification IV, words signifying abstractions, quality, states of mind, such as "platitudes," "torrid," "intellectuality," etc.

---

1 Elizabeth C. Morriss and Dorothy Holverson. "Idea Analysis Technique." Unpublished Manuscript, Teachers College, Columbia University, 1938.

Yoakam[1] in 1939 reported a technique based on an index difficulty derived from Thorndike's Teachers Word Book of 20,000 Words. Each word above the first four thousand most common words was given a number indicating the group to which it belonged; this was its index number. The page index was calculated for an average of ten selected pages.

Five of the above techniques were analyzed for reliability of measurement by Elliott[2] in 1941. The five methods were: Pressey-Lively, Patty-Painter, Yoakam, Vogel-Washburne, and Gray-Leary. She found little consistency in results when ranking the books rated by two formulas involving the same factors. The two formulas using vocabulary and sentence length gave twelve books out of twenty-eight the same position of difficulty, seven were ranked one position apart, six two ranks apart, and three ranked three positions apart.

The readability index calculated by Lorge[3] is a three factor regression formula: (1) average sentence

1 Gerald A. Yoakam. A Technique for Determining the Difficulty of Reading Material. Unpublished Material, University of Pittsburgh, 1939.

2 Catherine J. Elliott. "A Critical Analysis of the Objective Method of Measuring Reading Difficulty." Pittsburgh Schools, 25 (May-June 1941) 201-209.

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

length in words, (2) number of prepositional phrases, and (3) the number of different hard words. These three factors are also used by the Gray-Leary formula. A list of difficult words was constructed and checked against the Thorndike-Lorge Teachers Word Book of 30,000 Words and the Dale list. The McCall-Crabbs Standard Test Lessons in Reading were used as the criterion of comprehension.

The Marks of Readable Style was published by Flesch in 1943.[1] Using three elements of the Lorge Readability Index and two presumptive elements, Flesch analyzed 375 test pages of twenty-one magazines. These five factors were: (1) number of different words, (2) number of prepositional phrases, (3) average sentence length in words, (4) number of abstract words, and (5) number of affixed morphemes. The last three of these were used in his regression equation because they provided satisfactory differentiations from easy to highly difficult material for adults. Later the formula was revised,[2] using only the average sentence length in words and the number of syllables per one hundred words. The score obtained from this equation ranks the material from zero to one hundred,

1 Rudolf F. Flesch. Marks of Readable Style. Teachers College Contributions to Education, No. 897. Teachers College, Columbia University, 1943.

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

the higher the score the greater ease of reading. A "human interest" score was also obtained by using the number of personal sentences. The latter two formulas are discussed at length in Chapter III.

Dale and Chall[1] have done the most recent research in the field of readability reported here. They considered two factors in their multiple correlation equation: (1) the average sentence length in words and (2) what they called the "Dale Score," the number of words not appearing on the Dale list of three thousand most common words divided by the total number of words in the sample. The material analyzed was samples of one hundred words from every ten pages of health material. This formula will be elaborated upon in the following chapter.

Vocabulary

Many studies and lists have been made of vocabularies since vocabulary difficulty is a basic element in reading difficulty. The Bibliography of Vocabulary Studies by Dale[2] lists 1,855 references to research under twenty-five headings. An increase of 710 references is

---

1 Edgar Dale and Jeanne S. Chall. A Formula for Predicting Readability. Bureau of Educational Research, Ohio State University, 1948.

2 Edgar Dale. Bibliography of Vocabulary Studies. Bureau of Educational Research, Ohio State University, 1949.

listed by this bibliography over his 1939 edition.

Fries and Travers[1] give a complete history of word counts and limited vocabularies with notes on methods of choice, dictionary counts, inadequacy of mere frequency of use as a determiner of significance and the need for semantic counts.

Edward L. Thorndike has published three word lists. The first in 1921[2] is an alphabetical list of 10,000 most commonly occurring words in a count of 625,000 words from literature for children; 3,000,000 words from the Bible and English classics; 300,000 words from elementary school text books; 50,000 words from books about cooking, sewing, farming, the trades; 90,000 words from daily newspapers; and about 500,000 words from correspondence. Fifty-one different sources were covered in this count.

His second word count was A Teachers Wordbook of 20,000 Words.[3] In this later count over two hundred sources were covered. His original list was enlarged to

1   Charles C. Fries and A. Allen Travers. English Word
    Lists. Washington, D.C.: American Council on Edu-
    cation, 1940.

2   Edward L. Thorndike. The Teachers Word Book. New York:
    Bureau of Publications, Teachers College, Columbia
    University, 1921.

3   Edward L. Thorndike. A Teachers Wordbook of 20,000
    Words. New York: Bureau of Publications, Teachers
    College, Columbia University, 1931.

20,000 words and improvement was made on the selection of the most important 10,000 words appearing in the earlier study. The numbers of 1 to 20 in the Thorndike Junior Century Dictionary (1942), following the definitions, are in reference to the thousand into which words fall as to frequency of use, the smaller the number the more frequently the word is used.

Thorndike and Lorge combined their efforts to compile the Teachers Word Book of 30,000 Words.[1] Each word in this list was given a record of frequency of occurrence in general reading material and in four different sets of reading matter. A rating of 1 equaled at least one occurrence per million and not so many as two per million; a rating of 2 equaled at least two per million and not so many as three per million; and similarly up to 49. A rating of A equaled at least 50 per million; a rating of AA equaled 100 and over per million. Separate figures were given for the frequency of occurrence in the four and one-half million words of the Thorndike general count of 1931, the Lorge magazine count, the Thorndike count of 120 juvenile books and the Lorge-Thorndike semantic count.

Another project financed by grants from the Works Projects Administration was Rinsland's A Basic

---

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

Vocabulary of Elementary School Children.[1] Letters were
sent to 1,500 selected schools in all kinds of geographic,
economic, and social areas asking for samples of children's
writings, representing their freest and most natural
compositions....personal notes, stories, compositions in
many school subjects, examinations in non-technical
subjects, articles for school papers that were not corrected
by teachers, and reports on projects, trips, and obser-
vations,...100,212 compositions from the first eight grades
in the schools were used. Only one composition from each
child was accepted. There was found 25,632 different
words from the study of 6,012,359 running words, 14,571
words occurring three or more times in any one grade are
given in an alphabetical list. The actual or raw
frequency is given for occurrence in each of the eight
grades and the total for all grades. Also an index symbol
is given of the frequency group by the hundred, five
hundred, and thousand into which the word falls. The
total number of different words was found to be extremely
high, from 5,099 for the first grade to 17,930 for the
eighth grade.

Dale[2] compiled a list of 769 words common to
Thorndike's first thousand most frequent English words and

---

1 Henry D. Rinsland. A Basic Vocabulary of Elementary
   School Children. New York: Macmillian Company, 1947.

2 Edgar Dale. "A Comparison of Two Word Lists." Edu-
   cational Research Bulletin, 10 (December 1931) 484-498.

the first thousand most frequent words known to children entering the first grade. This last mentioned list was compiled by the Child Study Committee of the International Kindergarten Union. Dale noted, "one of the major problems confronting the writers of reading books for first grade students is the use of a vocabulary that satisfies two criteria: (1) that it be constituted of words known to children, and (2) that it contain those words which are likely to be of permanent value to children in their reading activities."

Gates[1] published a reading vocabulary for the primary grades. His original list was selected from: (1) the 2,500 words of highest frequency as determined by Thorndike's count, (2) any words not in the 2,500 from Thorndike found in the 1,000 words of highest frequency as determined by a count of words in a selection of children's literature, (3) all additional words in a series of readers for the primary grades, and (4) all additional words found in the thousand most frequent words in the spoken vocabulary of young children. Each of the 4,300 words was appraised for merit for use in reading at different stages during the primary grades on the basis of utility, interest, and difficulty by judgment of experts.

---

1 Arthur I. Gates. A Reading Vocabulary for Primary
   Grades. New York: Bureau of Publications, Teachers
   College, Columbia University, 1935.

Cole[1] directed a study in which the most widely used textbooks in each subject were first gone over by readers who were instructed to list all words occurring in the texts read which, because of being peculiar to the subject or uncommon, might be difficult for children in the grades in which the subject was taught. This list was then sent to teachers of that subject for them to check the words they considered: (1) absolutely essential, (2) important but not essential, or (3) unnecessary. Only four subjects were rated by less than thirty-five teachers.

Seegers and Seashore[2] startled the educational world in 1949. In an effort to throw some light on the size of children's meaning vocabularies, they summarized evidence from many sources. Interesting conclusions were reached which "raised serious question as to the accuracy of several very widely held educational beliefs, e.g., (1) that the initial vocabularies of school children are very small, (2) that the rate of growth of their vocabularies is proportionately small and difficult to improve without pushing the children, and (3) that it is therefore necessary to carefully control the nature and

---

1 Luella Cole. The Teachers Handbook of Technical Vocabulary. Bloomington, Illinois: Public School Publishing Company, 1940.

2 J. Conrad Seegers and R. H. Seashore. "How Large Are Children's Vocabularies?" Elementary English, 26 (April 1949) 181-194.

numbers of new terms presented by texts and curricular materials at each grade level." The investigators did not reconcile their conclusions with the objectively derived fact that current textbooks are loaded with unfamiliar words that children can secure the meaning of, if at all, only with great difficulty.

MacLatchy and Wardwell[1] determined the frequency of words in forty-two pre-primers. The total number of words used altogether was 1,929, a total of 289 different words was found and no one word was used in all forty-two pre-primers. The most common word was <u>and</u>, used in forty-one of the books. From this study seventy-one words were selected that "may be set up as the goal" during the pre-primer period.

In Table I, prepared by West,[2] are summaries of the size of the vocabulary of English-speaking children as estimated by various observers concerning children at various ages.

---

1   Josephine H. MacLatchy and Frances Wardwell. "Common Pre-Primer Words." <u>Educational Research Bulletin</u>, 26 (November 1948) 199-206, 226.

2   Michael West. <u>Bilingualism</u>. Bureau of Education, India, 1926.

TABLE I

THE SIZE OF VOCABULARY OF ENGLISH-SPEAKING
CHILDREN AS ESTIMATED BY VARIOUS
OBSERVERS

Age	Vocab.: No. of Words	Investigator	Age	Vocab.: No. of Words	Investigator
			1.3	235	Kirkpatrick[4]
2.0	215	Tracy[1]			
2.7	642	Salisbury[2]	2.8	405	"
5.5	1,528	"	3.8	700	"
6.5	2,500	Terman & Childs[3]			
7.5	2,600	"			
8.5	3,960	"	8.5	4,480	Kirkpatrick[5]
9.5	5,000	"	9.6	6,620	"
10.5	6,000	"	10.7	7,020	"
11.5	6,100	"	11.7	7,860	"
12.5	7,700	"	12.8	8,700	"
13.0	8,800	"	13.9	10,660	"
			15.0	12,000	"

1 F. Tracy. _Psychology_ _of_ _Childhood_. Boston: Ginn and Company, 1895.

2 A. Salisbury. _Educational_ _Review_, 7 (March 1894) 289-290.

3 L. M. Terman and H. G. Childs. _Journal_ _of_ _Educational_ _Psychology_, 3 (April 1912) 198-208.

4 E. A. Kirkpatrick, _Fundamentals_ _of_ _Child_ _Study_. New York: Teachers College, Columbia University, 1911.

5 E. A. Kirkpatrick. _Popular_ _Science_ _Monthly_, 70 (February 1921) 678.

Hartmann[1] found unexpectedly large recognition
vocabularies among college students. One word was
selected from the same relative position of every
fortieth page of the Merriam Webster's New International
Dictionary (unabridged lexicon). This total list of fifty
words represented 400,000 words, each sample word
representing approximately 8,000 words. This vocabulary
of fifty words was given to 106 students at the Alabama
Polytechnical Institute during the summer of 1945. In
order to have the recognition meaning of a word, any
inkling as to the meaning of the word was accepted, the
definitions did not have to be exact. From this study it
was estimated that the average college student has a
recognition vocabulary of 215,040 words.

An effective summary is given by Dolch[2] for the
use of vocabulary lists in predicting readability. When
undertaking to use such a list eight factors should be
considered: (1) vocabulary difficulty is a basic element
in reading difficulty, (2) vocabulary difficulty is only
one part of reading difficulty, (3) consider the problem

---

1 George W. Hartmann. "Further Evidence on the Unexpected
   Large Size of Recognition Vocabularies Among College
   Students." Journal of Educational Psychology, 27
   (October 1946) 436-439.

2 E. W. Dolch. "The Use of Vocabulary Lists in Predicting
   Readability and in Developing Reading Material."
   Elementary English, 26 (March 1949) 142.

of multiple meanings, (4) meaning vocabulary and sight
vocabulary are different things, (5) choose the size of
list to fit your needs, (6) study the words which are not
on the list, (7) consider the source of the list, and (8)
special subject matter lists must be considered in some
cases.

Reading Level

The nature and extent of the reading problem in
American education has been pointed out by Gray.[1] In a
study made in Chicago of 6,000 suburban ninth graders it
was found that twenty-two per cent of them read below the
seventh grade level, fifteen per cent below the sixth
grade level, and nine per cent below the fifth grade level.
Gray stated, "In my judgment the provision of appropriate
guidance in reading at these levels (high school and
college) is one of the most urgent needs of American
Education today."

Bond[2] found over ten per cent of the entering
ninth graders were three years or more retarded in

---

1 William S. Gray. "The Nature and Extent of the Reading
   Problems in American Education." Educational Records
   Supplement, (1938) 87-104.

2 Guy L. Bond. "Identifying the Reading Attainment and
   Needs of Students." Yearbook of the National Society
   for the Study of Education, (1948-2) 69-90.

vocabulary mastery. It was stated that this was an average high school. However, there were no students retarded more than one year academically, which seems far from a typical high school.

The reading grade levels were computed by Lorge and Blau[1] for adults by using the McCall-Thorndike Reading Scale. They found the average grade level for the fourteen year olds was 6.8, for twenty-five year olds it was 11.3, while for thirty-four year olds it was 9.2. This last figure agrees with that reported by Johnson[2] from unpublished Army and Navy data; the reading grade level for veterans of World War II was placed between grades eight and ten. In the article by Lorge and Blau, the authors stated, "The average adult, with his greater experience, with his matured interests and attitudes, will comprehend materials in a way quite superior to the average thirteen year old child.

Aukerman[3] found significant differences between the reading abilities of good and poor students in the eleventh grade paired in four academic subjects: history,

---

1   Irving Lorge and Ralph Blau. "Reading Comprehension of Adults." Teachers College Record, (1941) 189-198.

2   R. H. Johnson. "The Problem of Veteran's Reading Level in the Counseling Process." Minnesota Counselor, (1948) 3.

3   R. C. Aukerman, Jr. "Differences in the Reading Status of Good and Poor Eleventh Grade Students." Journal of Educational Research, 41 (March 1948) 498-515.

literature, science, and mathematics. Fourteen different
tests of general reading ability, specific reading ability,
specialized reading skills, general vocabulary ability, and
special vocabulary abilities were given and numerous
statistical comparisons were made. Reading and vocabulary
abilities by which good and poor students may be differ-
entiated varied from subject to subject. He also found
patterns of reading abilities existing in each of the
separate subjects. The results indicated that general
reading ability is the most significant differentiating
factor between good and poor eleventh grade students, when
paired, in all academic fields.

Research Relating Specifically to the Study

Few attempts have been made to judge the readability
of standardized testing instruments. Roeber[1] studied seven
of the most popular interest inventories. He was interested
only in the counting of words and checking the number of
words above the ninth grade level of difficulty. The
Thorndike list was used to determine the grade placement of
the words. The Occupational Interest Inventory-Advanced
(Lee-Thorpe) was the most wordy with 3,183 running words

1 Edward C. Roeber. "A Comparison of Seven Interest
   Inventories with Respect to Word Usage." Journal of
   Educational Research, 42 (September 1948) 8-17.

with 896 different words. However, 815 of these had an occurrence of five or more times per million, making the vocabulary a comparatively easy one. The Garretson-Symonds Interest Questionnaire for High School Students, 1942 edition, and the Thurstone Interest Schedule had the highest vocabulary difficulty. Both had slightly over twenty-one per cent of their words above the ninth grade level of understanding. However, the Thurstone used only 107 words and the Garretson-Symonds 512. The Strong Vocational Interest Blank (Men) had sixteen per cent of its 664 different words above the ninth grade of difficulty. The Cleeton Vocational Interest Inventory showed 784 words with fifteen per cent of them beyond the ninth grade level of understanding. The Kuder Preference Record (BB) used only 638 different words, even though it was rated next to the most wordy. Eleven per cent of the 638 words were above the ninth grade level.

On the basis of this study alone, it would seem questionable whether any one of these inventories is appropriate for the ninth grade.

The Lewerenz formula was applied by Stefflre[1] to six interest inventories. He justified the use of this formula on the basis that it is only concerned with

---

1 B. Stefflre. "The Reading Difficulty of Interest Inventories." Occupations, 26 (November 1947) 95-96.

vocabulary difficulty and not sentence structure. The
vocabulary difficulty grade placement was calculated for
each inventory: Occupational Preference Inventory-Form A
(Brainard), 6.4; Occupational Interest Inventory-Advanced
(Lee-Thorpe), 6.8; Kuder Preference Record-Form BB, 8.4;
Vocational Interest Blank for Men-Form A (Strong), 10.4;
A Study of Values (Allport-Vernon), 11.3; and the Cleeton
Vocational Interest Inventory-Form A, 12.0.

These findings were different from the previous
study although both were based on vocabulary difficulty.

Christensen[1] analyzed the words used in the Kuder
Preference Record and found that a large number of words
were used in it that were unfamiliar to the group being
tested. A test of twenty words was constructed from the
inventory and given to twenty-seven fourteen year olds of
average ability. Twenty-four per cent said that "A social
worker is a person who works for the success of socialism;"
twenty-six per cent responded yes to, "A certified public
accountant helps people find jobs most suited for them;"
thirty per cent said, "Sociology is the scientific study
of the habits of various animals;" and twenty-two per cent
marked plus the statement, "A psychologist draws maps for
geography books."

---

1 T. E. Christensen. "Some Observations with Respect to
   the Kuder Preference Record." Journal of Educational
   Research, 40 (October 1946) 96-107.

The average occurrence of the words per million was found to be three in the Kuder, suggesting that many of the words are too difficult for clear understanding of the items in which these words appear.

The most recent research that has come to the attention of the writer is that by Johnson and Bond.[1] Ten selected tests were subjected to the Flesch "Ease of Reading" formula. Table II shows the over-all readability levels of the selected tests as determined by application of the Flesch formula.

The authors stated that the Flesch and Lewerenz formulas showed relative general agreement on the readability level of the interest tests. From the above study by Stefflre it was seen that the interest "inventories" did appear in the same rank order; however, the Kuder being the most nearly in agreement was over one grade placement apart when appraised by the two formulas. The Allport-Vernon has a difference in grade placement of nearly four grades, and the Strong was scored 4.6 grades more difficult when rated by the Flesch formula.

---

1 Ralph H. Johnson and Guy L. Bond. "Reading Ease of Commonly Used Tests." _Journal of Applied Psychology_, 34 (October 1950) 319-324.

TABLE II

OVER-ALL READABILITY LEVELS OF SELECTED TESTS
AS DETERMINED BY APPLICATION OF
THE FLESCH FORMULA

Test	Reading Ease	Grade Level
Bennett Mechanical Comprehension	90	5.5
Minnesota Multiphasic Personality Inventory	88	5.5
Directions for Minnesota Clerical	87	6.0
Bell Adjustment Inventory	80	7.0
Directions for Bell	61	9.5*
California Interest Inventory	65	9.0*
Kuder Occupational Preference Record	60	9.5*
Directions for the Kuder	70	8.0
College G.E.D. No. 2	59	10.0*
Ohio State Psychological Test (Part 3)	37	15.0*
Strong Vocational Interest Blank	35	15.0*
Directions for the Strong	73	7.5
Allport-Vernon	35	15.0*
Directions for Allport-Vernon	60	9.5*

* Starred grade scores represent Flesch's corrected grade
placement for the area of extrapolation beyond the 7th
grade.

Summary

The review of the literature was divided into five

areas: annual summaries and bibliographies, vocabulary,

reading level, readability, and studies related specifically

to this study.

The large number of studies made in the field of reading in recent years necessitated careful selection of the studies to be included in this paper in order to prevent an endless review of the literature. The summaries and bibliographies were reviewed in order to give the reader a large amount of source material.

The section reviewing readability showed the lack of studies on measuring the reading difficulty of standardized tests, and that no method had been developed specifically for measuring objectively the readability of these tests.

The vocabulary studies were reviewed in order to show the importance of this factor on the readability of standardized tests. It was thought that this factor, vocabulary, would prove to be adequate for use in the proposed method for estimating the reading difficulty of standardized tests.

The reading level or attainment of children and adults in reading points out the seriousness of failing to consider the reading difficulty of printed material, if they are to understand what they read.

Few studies have been reported measuring the readability of standardized tests. The research that has been done used the formulas in existence for making their measurements rather than develop a specific measure for this purpose.

CHAPTER III

GUIDE TO THE WORKING OF THE VARIOUS FORMULAS

Introduction

In the previous chapter mention was made of several suggested methods for objective measurement of reading material. From these suggested methods the writer chose five for use in evaluating the reading difficulty of standardized testing instruments commonly used in counseling. These methods were: Dale-Chall, Flesch, Lorge, Lewerenz, and Yoakam.

A description of each of these methods is given in the following pages of this study.

The Dale-Chall Formula[1]

In view of the shortcomings of previous methods of evaluating objectively the readability of printed matter, Dale and Chall set up these hypotheses:

1. A large word list would predict as well as, if not better than, the count of affixes. It would avoid the pitfalls of lack of discrimination at the upper levels of difficulty.

---

1 Edgar Dale and Jeanne S. Chall. "A Formula for Predicting Readability." _Educational Research Bulletin_, 27 (January-February 1948) 11-20, 37-54.

2. A count of personal references does not add very much to the prediction of readability.

3. A shorter, more efficient formula could be evolved with the use of a word factor and a factor of sentence structure.

For their sample passages they used the McCall-Crabbs Standard Test Lessons in Reading,[1] the same passages used by Lorge and Flesch. These are a series of 376 passages of children's readings, already graded in difficulty on the basis of comprehensibility of questions at the end of each passage. This material, it should be noted, has serious deficiencies as a criterion, but is the best we have at the present time. The word count was based on the Dale list of approximately three thousand words. This was constructed several years ago by testing fourth graders on their knowledge in reading a list of approximately ten thousand words. This larger list included the most common words in the Thorndike,[2] Buckingham and Dolch[3] and other word lists. An attempt was made to include all words that fourth graders would possibly know. A word was considered

---

1 W. A. McCall and Lelah Crabbs. Standard Test Lessons in Reading. New York: Bureau of Publications, Teachers College, Columbia University, 1926.

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

3 B. R. Buckingham and E. W. Dolch. A Combined Word List. New York: Ginn and Company, 1936.

as known when at least eighty per cent of the fourth graders checked it as known.

The formula developed by Dale and Chall is based on vocabulary load, and a factor of sentence structure. Vocabulary load is the relative number of words outside of the Dale 3,000 words; the sentence structure factor on the average sentence length of the material.

Rules for selecting samples of a text to be analyzed and for computing the average sentence length and percentage of unfamiliar words is given in the following pages.

As each count is made, it is recorded on a work sheet where detailed steps are given for arriving at the grade level of reading difficulty.

Samples of approximately one hundred words about every tenth page is taken for books. For articles about four one hundred word samples per two thousand words should be selected. The samples are to be spaced evenly. For passages of about two hundred to three hundred words, the entire passage is to be analyzed. A sample should begin and end with complete sentences.

All words are to be counted in the sample. Hyphenated words and contractions are counted as single words. Numbers, such as 10 and 1947, are counted as single words. Compound names of Persons and places are single words. Initials which are part of the proper name are not

counted as separate words.

Complete sentences are counted.

Words that do not appear on the Dale list are considered unfamiliar. Each unfamiliar word is counted, even if they appear more than once.

The computation is simple after these counts have been made. The average sentence length is computed by dividing the number of words in the sample by the number of sentences in the sample. The Dale score or percentage of words outside of the Dale list is computed by dividing the number of words not on the Dale list by the number of words in the sample, and multiplying by one hundred. The average sentence length is multiplied by .0496 and the Dale score by .1579. To these is added the constant 3.6365, this giving the formula raw score. If more than one sample is analyzed, the average of the formula raw scores is obtained. The average raw score is converted into a corrected grade level score from the following correction table:

Formula Raw Score	Corrected Grade Level
4.9 and below	4th grade and below
5.0 to 5.9	5-6th grade
6.0 to 6.9	7-8th grade
7.0 to 7.9	9-10th grade
8.0 to 8.9	11-12th grade
9.0 to 9.9	13-15th grade
10.0 and above	16/ (college graduate)

The corrected grade level obtained from the above table is the grade at which a book or article can be understood. For example, reading material which is figured at a corrected grade level of 11-12 is reading that should be within the ability of the average junior or senior in high school. Should the material be selected for adults, this corrected grade level of 11-12 means that a person who has had three or four years of high school ought to be able to read the material with ease and understanding.

The counting of the unfamiliar words is the difficult part of the formula. The authors give a long list of "special rules" that they state are necessary for the tabulation of common proper nouns, verbs, and other parts of speech. One should be well versed in grammar in order to apply many of these "special rules."

The Flesch Formula[1]

This revised formula by Flesch is much less time consuming than the earlier one published in 1943.[2] In the earlier readability formula, written as a doctoral

---

1 Rudolf R. Flesch. "A New Readability Yardstick." _Journal of Applied Psychology_, 32 (June 1948) 221-233.

2 Rudolf F. Flesch. _Marks of Readable Style_. Contributions to Education, No. 897. Teachers College, Columbia University, 1943.

dissertation, he counted the number of affixed morphemes in addition to average sentence length in words. The affixed morphemes are elements of language expressing connection between ideas that are attached to the semantemes which are linguistic elements expressing ideas or concepts. Morphemes express the relations established in the mind between semantemes; phonetic elements joined to the semantemes are affixed morphemes.

For a more salable book and a formula that could be understood by lay people, Flesch dropped the counting of affixed morphemes and counted syllables instead. His formula is a two-factor one: (1) the average sentence length in words and (2) the number of syllables per one hundred words.

In taking the samples for working this formula, it is suggested that three to five samples be taken for articles and twenty-five to thirty samples be taken for books. The samples are to be one hundred words in length, each starting at the beginning of a paragraph.

Count as words all letters, numbers, or symbols, or count as words the groups of letters, numbers, or symbols that are surrounded by white space. Count contractions and hyphenated words as one word. For example, count each of the following as one word: 1948, $19,892, e.g., C.O.D., wouldn't, and week-end.

As the next step, figure the average number of words in the sentences. Count the number of sentences in all the samples and divide by the total number of words in the samples. The last sentence in the sample should be counted if more than half of the words in this last sentence fall within the one hundred word sample selected.

Figure the average word length in syllables per hundred words. To do this count all syllables and divide the total number of syllables by the number of words. In the formula, this measure is expressed as the number of syllables per hundred words; therefore, multiply by 100.

Multiply the average sentence length by 1.015, the number of syllables per hundred words by .846; add the results of these two computations and from it subtract the constant 206.835. This becomes the Reading Ease Score. These scores range from 0 to 100; as the scores approach 0, the more difficult the reading becomes, and as the score moves toward 100 the easier it gets. The following table gives the relative difficulty of the reading material and where such writing would most commonly be found for the various Reading Ease Scores.

Reading Ease Score	Description of Style	Typical Magazine
90-100	Very Easy	Comics
80- 90	Easy	Pulp fiction
70- 80	Fairly Easy	Slick fiction
60- 70	Standard	Digests, _Time_
50- 60	Fairly Difficult	_Harpers_, _Atlantic_
30- 50	Difficult	Academic, Scholarly
0- 30	Very Difficult	Scientific, Professional

The four other formulas used in this study express readability scores in grade levels. Flesch[1] has translated the scores for his readability formula into grade level scores:

Ease of Reading Score	Grade Level
90-100	5th grade
80- 90	6th grade
70- 80	7th grade
60- 70	8-9th grade
50- 60	10-12th grade (high school)
30- 50	13-16th grade (college)
0- 30	College Graduate

---

1 Rudolf F. Flesch. _How to Test Readability_. New York: Harper and Brothers, 1951.

The Lorge Formula[1]

The Lorge formula is a means of judging the relative difficulty of either read or spoken passages. The readability index is an estimate of the reading grade at which the average school child will be able to answer about fifty per cent of the questions concerning detail, appreciation, import, vocabulary, and concept with adequate completeness and correctness. The Lorge formula is based on a criterion derived from responses to questions of these five types. It tends, therefore, to over-estimate the difficulty of passages to be read primarily for appreciation or for general import; it tends to under-estimate the difficulty of passages to be read primarily for specific details or for following directions. Nevertheless, the Lorge formula provides an over-all estimate which should be useful in grading materials. As an estimate, it should not be considered definitive, nor should it be used blindly. The readability index of Lorge is an estimate or approximation, and not a rigorous determination.

When a short passage is to be appraised, it is advisable to analyze the entire passage. Longer passages may be appraised by taking samples from the material.

---

1 Irving Lorge. <u>Predicting Readability</u>. Teachers College Record, No. 6, 1944.

Select a sample near the beginning, another near the
middle, and a third sample near the end of the passage.
These samples should be approximately one hundred words
in length. If the passage is less than three hundred words
in length, the entire passage should be analyzed. A good
procedure might be to number the lines of a lengthy passage
serially, then count the number of words per line (for about
ten lines) to get an estimate of the number of words. From
these numbered lines choose the number necessary in order
to give the size of sample desired. A sample should start
with the beginning of a sentence, and should stop at the
end of a sentence.

When books are to be appraised, it is advisable
to analyze from five to ten per cent of the book, never
less than five samples. For a book of ninety pages, with
an average of 195 words per page, this would mean 900
words for a five per cent sample and 1,800 words for the
ten per cent sample. It is better to take these samples
by the page. The five per cent sample would be every
eighteenth page used for analysis; every tenth page
should be chosen for the ten per cent sample.

Each word is counted in the sample. Hyphenated
words are counted as one word. When in doubt about uncommon
hyphenations, follow Webster's Unabridged Dictionary (2nd
edition); if listed in the dictionary as hyphenated, count

as one word; if not listed, count as two words. Words
separated at the end of a line are counted as a single
word. Numbers are counted as words; 1940 is a single
word, nineteen-forty. Compound words like place names
and person's names are counted as one word, e.g., New York,
United States, van Loon, Santa Claus, St. Nicholas.
Contractions are counted as one word.

Begin at the beginning of the sample and count
the number of complete sentences.

Count each prepositional phrase in the sample.
A phrase is made up of a preposition and a noun, a
preposition and a pronoun, or a preposition and a gerund,
e.g., to the house (noun), for him (pronoun), in skating
(gerund). Some less common prepositions are: despite,
concerning, notwithstanding. Infinitive phrases are not
to be counted. An infinitive phrase is made up of a
preposition (to) and a verb, e.g., to swim, to answer.
If a prepositional word is part of a clause, it is a
conjunction, and hence is not counted, e.g., "After the
storm had passed" is not counted. Some knowledge of
grammar is necessary in order to apply this part of the
Lorge formula.

Use the Dale list to cross out in the sample
every word on the Dale list, regardless of meaning. This
count is the number of different hard words, so that each
hard word is counted only once. Many special rules are

given for the counting of hard words.

Computation is rather simple: divide the number of words in the sample by the number of sentences; this is multiplied by .06. Divide the number of prepositional phrases by the number of words in the sample; multiply by 9.55. Divide the number of hard words by the total number of words; multiply by 10.43. These three figures plus the constant 1.9892 are added to secure the <u>Readability</u> <u>Index</u>.

A readability index of 9.2 for a passage may be considered indicative of the reading material at the ninth grade; it may be thought of in terms of placement of the material as within the reading comprehension of the average ninth grade child. The Lorge formula, in addition to its use for estimating the reading difficulty of passages for grade placement, may be used to advantage in estimating the difficulty of reading and of oral passages for adults. The formula yields a readability index which places materials in relative order of difficulty. A reading passage with an index of 6.2 may be considered less diffi-cult than a reading passage with an index of 8.1, etc.

The Lewerenz Formula[1]

The difficulty of a given selection of reading matter may be occasioned by several factors of which vocabulary is usually one. The same thought may be expressed in simple or difficult words, according to the style of the author. A given selection may be easy or hard, depending on the reader's own vocabulary comprehension.

The purpose of the Lewerenz formula described below is to compare a given book or article with a known scale of vocabulary difficulty. Six uses for a vocabulary grade placement formula may be pointed out:

1. Selection of textbooks which will be appropriate to the grade in which they will be used.

2. Selection of readers which in difficulty will fit the needs of specific grade groups.

3. Selection of supplementary reading materials for dull over-age pupils and young pupils of superior mentality. The latter group usually needs stories with a juvenile theme but written with a somewhat mature style, while the former group enjoys books with a theme suited to its social maturity but told in a very simple manner.

---

1 Alfred S. Lewerenz. "Measurement of the Difficulty of Reading Materials." Educational Research Bulletin Los Angeles Public Schools, 8 (March 1939) 11-16.

4. Measurement of the compositions of children to form a basis for estimating their knowledge of words.

5. Checking for difficulty of vocabularies used in spelling lists in order to keep them balanced and within the grade level where used.

6. Enabling librarians to recommend books more effectively to teachers who wish to supply their children with recreational reading materials fitted to their comprehension level.

Lewerenz set out to simplify the "Winnetka Formula." In the case of the "Winnetka Method," words were recorded on little slips of paperk one by three inches in size. When an evaluator alphabetized these slips they frequently would be blown off the table by the sudden opening of a door or by a gust of wind from the window. A great deal of time was spent with each book, writing down the common words. A large sheet of paper is used by Lewerenz to off-set these two objections. Up to this time measures of readability had been planned for use mainly with recreational reading. Lewerenz wanted to produce a measure which would yield results in terms of threshold reading ability. As has been pointed out in the previous formulas, a great amount of knowledge about grammar is necessary in order to apply the formulas. Lewerenz has worked out a method which requires little technical preparation in grammar and English construction.

Several methods of word counting were carried on before it was noticed that the words beginning with certain letters became increasingly more difficult as the lists of words were lengthened. The letters b, h, and w were initial letters of easy words; e and i were the only two initial letters that indicated difficult words.

Validity of the vocabulary difficulty was tested in several ways. The fact that the index letter values are based upon the norms of the Stanford Achievement Test, Reading Examination, makes them in accord with the performance of a great many children on the test. The formula has been applied to word lists designed for use in the primary grades and the grade placements have compared favorably with the grades for which the lists were designed. The formula has been applied to six different series of graded reading books. The results have shown that while the reading tests may get increasingly difficult from a vocabulary standpoint, they do not do so as rapidly as the grade indicated on the cover would imply. The validity of the formula shows consistency of grade placement derived from books which have been carefully graded and from the V and W forms of the Stanford Achievement Test which were used to recheck the original norms.

The following steps should be taken in finding the vocabulary difficulty of a given selection of reading matter.

The number of different words in one thousand words is found by taking a sampling from the right end of the third line (either top or bottom) of approximately all printed pages of the material being evaluated. For example, if a book has 250 pages, 4 words from the third line on every page would need to be taken to yield the necessary 1,000 words.

Proper nouns are included as regular words. Hyphenated words are counted as two words where each can stand alone in the sense in which used. For example, "heart-free" is counted as two words, but "Pan-American" is counted as one word. Derived forms of a given word are not counted as separate words. In case of doubt whether to record a word as different from some form already tabulated, Webster's Dictionary, preferably the unabridged edition, should be consulted.

After the samplings are made, the number of different words beginning with each letter is tabulated. This is the grand total. The totals for b, e, h, i, and w words are entered in the formula. Each of these totals is divided by the grand total of different words which gives the per cent of each. This figure should be carried to the nearest hundredth per cent. The values for the key-letter words in per cents are found by consulting the "Table of Converting Index Letter Per Cents into Values,"

Table III. After the values for all five of the index
letters are found, they are totaled and multiplied by .02.
The product is the _Vocabulary_ _Difficulty_ _Grade_ _Placement_
Score.

Thus, if the values for the five letters are,
respectively, 100, 95, 110, 81, and 98, the total will be
484. The sum multiplied by .02 gives a vocabulary
difficulty grade placement of 9.68. An average child in
the ninth grade should be able to read such material with
ease and understanding of the words involved. An adult
who has progressed into the ninth grade before leaving
school should also understand the vernacular of this read-
ing material.

TABLE III

TABLE FOR CONVERTING INDEX LETTER PER CENTS INTO VALUES

Index Letter Value	B%	E%	H%	I%	W%	Index Letter Value	B%	E%	H%	I%	W%
15	7.50	1.60	7.20	2.10	8.10	73	5.26	4.10	3.76	4.00	5.40
16	7.48	1.63	7.15	2.12	8.05	74	5.23	4.15	3.73	4.05	5.35
17	7.46	1.66	7.10	2.14	8.00	75	5.20	4.20	3.70	4.10	5.30
18	7.44	1.70	7.00	2.16	7.96	76	5.15	4.25	3.68	4.13	5.25
19	7.42	1.75	6.95	2.18	7.93	77	5.10	4.30	3.66	4.16	5.20
20	7.40	1.80	6.90	2.20	7.90	78	6.05	4.40	3.64	4.20	5.16
21	7.39	1.83	6.85	2.21	7.85	79	5.03	4.45	3.62	4.25	5.13
22	7.38	1.86	6.80	2.22	7.80	80	5.00	4.50	3.60	4.30	5.10
23	7.37	1.90	6.70	2.23	7.76	81	4.98	4.53	3.58	4.33	5.05
24	7.36	1.95	6.65	2.24	7.73	82	4.96	4.56	3.56	4.36	5.00
25	7.35	2.00	6.60	2.25	7.70	83	4.94	4.60	3.54	4.40	4.96
26	7.34	2.05	6.50	2.26	7.60	84	4.92	4.65	3.52	4.45	4.93
27	7.33	2.10	6.40	2.27	7.50	85	4.90	4.70	3.50	4.50	4.90
28	7.32	2.20	6.30	2.28	7.40	86	4.88	4.72	3.45	4.52	4.85
29	7.31	2.25	6.20	2.29	7.35	87	4.86	4.74	3.40	4.54	4.80
30	7.30	2.30	6.10	2.30	7.30	88	4.84	4.76	3.39	4.56	4.76
31	7.25	2.35	6.00	2.32	7.25	89	4.82	4.78	3.38	4.58	4.73
32	7.20	2.40	5.90	2.34	7.20	90	4.80	4.80	3.37	4.60	4.70
33	7.16	2.50	5.80	2.36	7.16	91	4.78	4.81	3.35	4.63	4.65
34	7.13	2.60	5.75	2.38	7.13	92	4.76	4.82	3.33	4.66	4.60
35	7.10	2.70	5.70	2.40	7.10	93	4.74	4.83	3.32	4.70	4.50
36	7.05	2.75	5.60	2.45	7.08	94	4.72	4.84	3.31	3.75	4.45
37	7.00	2.80	5.50	2.50	7.06	95	4.70	4.85	3.30	4.80	4.40
38	6.90	2.90	5.40	2.60	7.04	96	4.65	4.86	3.29	4.81	4.35
39	6.85	2.95	5.30	2.65	7.02	97	4.60	4.87	3.28	4.82	4.30
40	6.80	3.00	5.29	2.70	7.00	98	4.56	4.88	3.27	4.83	4.26
41	6.70	3.05	5.27	2.73	6.95	99	4.53	4.89	3.26	4.84	4.23
42	6.60	3.10	5.25	2.76	6.90	100	4.50	4.90	3.25	4.85	4.20
43	6.50	3.20	5.23	2.80	6.86	101	4.48	4.92	3.24	4.86	4.18
44	6.45	3.25	5.21	2.85	6.83	102	4.46	4.94	3.23	4.87	4.16
45	6.40	3.30	5.20	2.90	6.80	103	4.44	4.96	3.22	4.88	4.14
46	6.30	3.33	5.15	2.93	6.75	104	4.42	4.98	3.21	4.89	4.12
47	6.20	3.36	5.10	2.96	6.70	105	4.40	5.00	3.20	4.90	4.10
48	6.10	3.40	5.06	3.00	6.66	106	4.35	5.03	3.19	4.91	4.09
49	6.05	3.45	5.03	3.05	6.63	107	4.30	5.06	3.18	4.92	4.08
50	6.00	3.50	5.00	3.10	6.60	108	4.20	5.10	3.17	4.93	4.07
51	5.95	3.53	4.95	3.13	6.55	109	4.15	5.15	3.16	4.94	4.06
52	5.90	3.56	4.90	3.16	6.50	110	4.10	5.20	3.15	4.95	4.05
53	5.86	3.60	4.80	3.20	6.46	111	4.05	5.22	3.14	4.96	4.04
54	5.83	3.65	4.75	3.25	6.43	112	4.00	5.23	3.13	4.97	4.03
55	5.80	3.70	4.70	3.30	6.40	113	3.96	5.24	3.12	4.98	4.02
56	5.75	3.72	4.65	3.33	6.35	114	3.93	5.25	3.11	4.99	4.01
57	5.70	3.74	4.60	3.36	6.30	115	3.90	5.26	3.10	5.00	4.00
58	5.66	3.76	4.50	3.40	6.26	116	3.85	5.27	3.09	5.01	3.99
59	5.63	3.78	4.45	3.45	6.23	117	3.80	5.29	3.08	5.02	3.98
60	5.60	3.80	4.40	3.50	6.20	118	3.75	5.30	3.07	5.03	3.97
61	5.58	3.82	4.35	3.53	6.10	119	3.90	5.35	3.06	5.04	3.96
62	5.56	3.84	4.30	3.56	6.05	220	3.60	5.40	3.05	5.05	3.95
63	5.54	3.86	4.26	3.60	6.00	121	3.55	5.45	3.04	5.07	3.93
64	5.52	3.88	4.23	3.65	5.95	122	3.50	5.50	3.03	5.07	3.93
65	5.50	3.90	4.20	3.70	5.90	123	3.46	5.55	3.02	5.08	3.92
66	5.48	3.92	4.15	3.73	5.85	124	3.43	5.60	3.01	5.09	3.91
67	5.46	3.94	4.10	3.76	5.80	125	3.40	5.63	3.00	5.10	3.90
68	5.44	3.96	4.00	3.80	5.76	126	3.36	4.66	2.98	5.13	3.88
69	5.42	3.98	3.95	3.85	5.73	127	3.33	5.70	2.95	5.15	3.85
70	5.40	4.00	3.90	3.90	5.70	128	3.30	5.75	2.93	5.18	3.83
71	5.35	4.03	3.85	3.93	5.60	129	3.20	5.80	2.90	5.20	3.80
72	5.30	4.06	3.80	3.96	5.50						

The Yoakam Formula[1]

The technique for determining the reading level
of books developed by Yoakam is simple to use and economical
of time. In accuracy it compares favorably with longer and
more time-consuming techniques developed by others. With
practice the grader can find the grade level of a book in
one hour or less.

For practical purposes Yoakam says a sampling of
ten _selected_ pages distributed at approximately equal
intervals throughout the book is reasonably reliable.
Fifteen selected pages is better, but increases the amount
of time by half. A ten page sampling will ordinarily bring
one within .6 of a grade of the score secured by applying
the formula to the entire book.

The technique may be used for sampling books
from the fourth grade through high school and also yields
apparently reliable results for general literature. It
cannot be used for sampling books in mathematics or
subjects where formulas or mathematical symbols constitute
a large proportion of the content, but may be used to
sample textual matter where the bulk of the matter con-
sists of connected discourse.

---

1 Gerald A. Yoakam. _Readability Formula._ Unpublished
   Manuscript. University of Pittsburgh, 1948.

A preliminary survey of the book should be made in order to get an idea of its nature and to determine the typical amount of textual matter on different pages. If a book consists entirely of full pages of textual matter without illustrations, the sampling is comparatively easy; simply take ten full pages distributed at approximately equal distances throughout the book. Should the book vary greatly as to the page make-up, select the ten-page sample so as to get a some-what typical picture of the book.

The average number of running words on a full page should be computed by either counting three full pages and averaging, or by estimating the size of three full pages and averaging the estimates. Preface, introduction, bibliographies and test exercises in textbooks should not be included in the samplings.

Scan the sampled pages and underline in pencil each word which appears difficult, or if preferred, list all scored words from the page with their index numbers. These index numbers can be found in the Thorndike Teachers Wordbook of Twenty Thousand Words. Write the index number of each work of four or above, over or opposite the word underlined. All words on the page will have been indexed save those bearing the index numbers of 1, 2, or 3 in the Thorndike list.

With practice one is soon able to identify all words of above the fourth thousand quite easily. The words

of four and above constitute only a small proportion of the total running words on a page. Ordinarily there is seldom more than twenty-five on a page, depending upon the difficulty of the material. Score each word only once on each page, and in each chapter or division.

Score all compound words which are not hyphenated by using the Thorndike index number of that word even though it seems high. Score a compound word containing a hyphen by averaging the separate index numbers of the two parts, unless the word is scored by Thorndike, in which case give it the Thorndike index number.

Give all the words that do not appear in the Thorndike List a value of 20.

Proper names are not scored. Yoakam is carrying on at the present time a study to determine the effect of proper names on difficulty of reading matter.

Give all variants of a root word the same index number as the root, unless Thorndike scores the variant. Expletives or nonsense words are not scored.

Compute the page index number by adding the index numbers of all words with an index number of 4 or above found on the sampled page. Add the index numbers of the ten different pages and determine the average page index number, this being the book index number.

In order to determine the grade placement of a book one must consult the Reading Difficulty Scale. This

scale is found in Table IV. Taking the book index number
and the average number of words per page the grade place-
ment is read directly from this table. The grades range
from three to fourteen, and interpolation is possible,
placing the book within the tenth of a grade placement
score.

The grade levels of this formula are as yet
tentative. If books are properly sampled, the grade level
scores will reveal the differences between books as
determined by the basic difficulty of vocabulary, which is
the most important general element of difficulty. The
grade placement of a book classifies the book into the
grade level where it should be read with little difficulty
by students with average reading ability for that grade.

TABLE IV

READING DIFFICULTY SCALE

Grade	100	D	120	D	140	D	160	D	180	D	200	D	220	D	240	D
3	3	7	4	8	5	9	6	10	7	11	8	12	9	13	10	15
4	10	7	12	8	14	9	16	11	18	11	20	13	22	14	25	15
5	17	7	20	9	23	10	27	11	30	12	33	14	36	15	40	16
6	24	8	29	9	33	11	30	12	42	13	47	14	51	16	56	17
7	32	8	38	10	44	12	50	13	55	15	61	16	67	17	73	18
8	40	9	48	11	56	12	63	14	70	16	77	17	84	18	91	20
9	49	9	59	11	68	13	77	15	86	17	94	18	102	19	111	21
10	58	9	70	11	81	13	92	15	103	17	112	19	121	20	132	22
11	67	10	81	11	94	13	107	15	120	17	131	20	141	21	154	24
12	77	10	92	11	107	13	122	15	137	17	151	21	162	22	178	24
13	87	11	103	11	120	13	137	15	154	18	172	21	184	23	202	25
14	98	11	114	11	133	13	152	15	172	18	193	21	207	23	237	25

Grade	260	D	280	D	300	D	320	D	340	D	360	D	380	D	400	D
3	11	16	12	17	13	18	14	19	16	19	17	20	18	22	19	24
4	27	16	29	17	31	19	33	20	35	20	37	22	40	24	43	26
5	43	17	46	18	50	20	53	22	55	22	59	24	64	24	69	26
6	60	18	64	19	79	21	75	24	77	24	83	26	88	26	95	28
7	78	19	83	21	91	22	99	26	101	26	109	28	114	28	124	30
8	97	21	104	22	113	24	125	28	127	28	137	30	142	30	154	32
9	118	23	126	24	137	26	153	30	155	30	167	32	172	32	186	34
10	141	25	150	26	163	28	183	31	185	31	199	34	204	34	220	36
11	166	26	176	28	199	30	214	32	221	32	233	36	238	36	256	38
12	192	27	204	30	229	31	246	32	253	32	269	38	274	38	295	30
13	219	27	234	30	260	31	278	32	285	34	307	40	312	40	335	42

Comparative Analysis of the Five Methods

Before discussing the application of the five techniques for judging readability, it will be of interest to make a brief comparison of them for the purpose of evaluating standardized tests.

The Lorge formula uses three variables in determining the inherent reading difficulty of various materials. Lorge bases his predictions on the factors of vocabulary, average sentence length, and relative number of prepositional phrases. Two of the formulas, the Dale-Chall and the Flesch, use two variables each. The Dale-Chall formula uses only the first two Lorge factors, average sentence length and vocabulary difficulty. The Flesch formula involves the average sentence length and the average number of syllables per 100 words. The Lewerenz and the Yoakam formulas rely upon but a single factor to determine readability. For the Lewerenz formula a count of the number of words beginning with b, h, and w is made for the easy words, and the words beginning with the letters e and i are counted for the difficult words. Each word beginning with one of the five letters is given a value which influences the reading difficulty grade placement score. In the Yoakam formula each word above the most common four thousand is given a weight depending on the commonness of the word. The total of these weights gives the reading difficulty score.

Table V gives a condensed picture of the factors involved in each of the five formulas.

TABLE V

READABILITY FACTORS INVOLVED IN EACH OF THE FIVE FORMULAS

Factor	Dale-Chall	Flesch	Lorge	Lewerenz	Yoakam
Average Sentence Length	X	X	X		
Difficult Words	X		X		X
Number of Syllables		X			
Prepositional Phrases			X		
Different Words				X	
Initial Letters				X	

Word difficulty is basic to all reading materials and particularly to reading involved in testing instruments, in which there are often only clauses, phrases or words to read and understand.

Each of the five formulas interpret word difficulty in a different light. The Lorge formula considers any word other than the 769 words that are common to the first one thousand most frequent English words on the Thorndike list and the first thousand most frequent words known by children entering the first grade as a "hand word." The Dale-Chall

formula classes any word above the Dale list of three thousand as difficult. To the Yoakam formula, "hard words" vary in difficulty according to their frequency and range of occurrence above the most common four thousand words. The Lewerenz formula is based solely on word difficulty, basing the vocabulary difficulty on words with certain initial letters. The Flesch formula considers the length of the word the index of difficulty of that word, the more syllables a word has the more difficult it is.

Without any consideration to influence the opinion other than the facts mentioned above, it would seem that the Lorge formula interpretation of "hard words" is too simple and limited; the Dale-Chall method approaches a more realistic and practical definition of difficult words; and the Yoakam formula is perhaps the most realistic of all the formulas for use with testing instruments. The idea that difficult and easy words begin with certain letters, presented in the Lewerenz formula, does not seem to apply to standardized tests; and the number of syllables that a word has, as proposed by Flesch, does not necessarily give its index of difficulty.

Summary

The five formulas selected for study are the more recently developed techniques for measuring readability.

These formulas present several factors which have been
used for determining reading difficulty of printed
matter, namely, word difficulty, prepositional phrases,
sentence length, number of syllables per one hundred
words, number of different words, and percentage of words
beginning with certain letters. Each of the formulas has
been carefully developed and exhibits a fair degree of
reliability and validity.

CHAPTER IV

METHODS OF ATTACK

Selection of Tests

The primary purpose of the present study was to find an acceptable method for grading the readability of standardized tests. Formulas in current use do not lend themselves to the appraising of testing instruments. The general plan of this study was to apply the more recent formulas to twenty-seven commonly used standardized tests in order to develop a more adequate method of determining their readability.

The choice of the tests to be used in this study was determined from previous studies made upon test preference and from newer tests indicated by the records of the University of Kansas Guidance Bureau.

Berkshire, et al.[1] have tabulated responses that were received from 290 testing centers. They concluded that there is general agreement on approximately 15 to 20 tests as being common to guidance testing. Beyond this point test preference varies widely. Tests were chosen from this list for analysis, if they were reported by at least twenty-five of the reporting centers as being one of

---

1 J. R. Berkshire, J. E. T. Bugental, Frank P. Cassens, and Harold A. Edgerton. "Test Preference in Guidance Centers." Occupations, 26 (March 1948) 337-343.

the ten most commonly used tests.

This same study shows in tabular form the results obtained from three other studies by Brophy and Long,[1] Darley and Marquis,[2] and Baker and Peatman.[3] The findings from these three studies were comparable to those of Berkshire and others.

Standardized testing instruments that have become very popular since the appearance of the above articles were checked for the frequency of use at the University of Kansas Guidance Bureau. Nine tests were added to the list to be analyzed. Six of these that were added to the list for analysis were published after the above cited studies on preference had been made. Two of the remaining three that were added were reading tests, because of the nature of the study--readability. The one remaining test, Minnesota Personality Scale (Men) 1941, is the test for which there was no apparent reason for omission from the studies of test preference.

---

1 D. F. Brophy and L. Long. "Veterans Administration Vocational Training Program: Processing Procedures Used by the College of the City of New York." Psychological Bulletin, 41 (November 1944) 795-802.

2 J. G. Darley and D. G. Marquis. "Veterans' Guidance Centers: A Survey of Their Problems and Activities." Journal of Clinical Psychology, 2 (January 1946) 109-116.

3 G. Baker and J. G. Peatman. "Tests Used in Veterans Administration Advisement Units." American Psychologist, 2 (March 1947) 99-102.

The tests were chosen from five of the general areas of testing listed by Buros:[1] Character and Personality, Intelligence (group), Interests, Achievement Batteries, and Reading.

Character and Personality Tests

1. The Adjustment Inventory--Student Form (grades 9-16), Hugh M. Bell, 1934, Stanford University Press.

2. California Test of Personality--Adult, Form A, Ernest W. Tiegs, Willis W. Clark, and Louis P. Thorpe, 1942, California Test Bureau.

3. The Guilford-Zimmerman Temperament Survey, (grades 9-16 and adults), 1949, Sheridan Supply Company.

4. Minnesota Multiphasic Personality Inventory, (ages 16 and over), Starke R. Hathaway and J. Charnley McKinley, 1943, Psychological Corporation, New York.

5. Minnesota Personality Scale--Men (grades 11-16), John G. Darley and Walter J. McNamara, 1941, Psychological Corporation, New York.

6. Mooney Problem Check List--High School, Form H (grades 9-12), Ross L. Mooney, 1950 Revision, Bureau of Educational Research, Ohio State University.

7. The Personality Inventory (grades 9-16 and adults), Robert G. Bernreuter, 1935, Stanford University Press.

---

1 Oscar Krisen Buros, editor. The Third Mental Measurements Yearbook. New Brunswick: Rutgers University Press. 1949.

8. <u>The School Inventory</u> (grades 9-12), Hugh
M. Bell, 1948, Stanford University Press.

9. <u>A Study of Values</u> (college and adult),
Gordon W. Allport and Philip E. Vernon, 1931, Houghton
Mifflin Co.

Group Intelligence Tests

1. <u>American Council on Education Psychological
Examination</u> (college freshmen), L. L. Thurstone and Thelma
Gwinn Thurstone, 1948, Educational Testing Service.

2. <u>Army General Classification Test--First
Civilian Edition--</u>Form AM (grades 9-16 and adults), 1947,
Science Research Associates.

3. <u>California Test of Mental Maturity--</u>Advanced
Series (grades 9-adults), Elizabeth T. Sullivian, Willis
W. Clark, and Ernest W. Tiegs, 1946 revision, California
Test Bureau.

4. <u>The Henmon-Nelson Test of Mental Ability--</u>
Form A (Intelligence Test for College Students), V. A. C.
Henmon and N. J. Nelson, 1931, Houghton Mifflin Company.

5. <u>Ohio State University Psychological Test--</u>
Form 22, (grades 9-16 and adults), Herbert A. Toops, 1943,
Bureau of Educational Research, Ohio State University.

6. <u>Otis Employment Tests--</u>Form 2A (formerly
called Otis Self-Administering Test of Mental Ability),
(grades 9-16 and adults), Arthur S. Otis, 1943, World Book
Company.

7. Otis Quick-Scoring Mental Ability Tests--
Form Gamma AM (for high schools and colleges), Arthur S.
Otis, 1937, World Book Company.

Tests of Interests

1. Cleeton Vocational Interest Inventory (Men)--
Form A, (grades 9-16 and adults), Glen U. Cleeton, 1943,
McKnight and McKnight, Publishers.

2. Kuder Preference Record--Form BB, (grades 9-16
and adults), G. Frederic Kuder, 1942, Science Research
Associates.

3. Kuder Preference Record-Vocational--Form CM,
(grades 9-16 and adults), G. Frederic Kuder, 1948, Science
Research Associates.

4. Occupational Interest Inventory--Advanced,
Form A, (grades 10-13 and adults), Edwin A. Lee and Louis
P. Thorpe, 1943, California Test Bureau.

5. Vocational Interest Blank for Men (Revised)--
Form MM, (ages 17 and over), Edward K. Strong, Jr., 1938,
Stanford University Press.

Achievement Test Batteries

1. Cooperative General Culture Test--Form XX,
(a revision of form X, College), Norman J. Blair, Jeanne M.
Bradford, Miriam M. Bryan, Paul J. Burke, and Herbert
Danzer, 1951, Educational Testing Service.

2. <u>Stanford Achievement Test</u>--Advanced Battery--Complete Form D, (grades 7-9), Truman L. Kelley, Giles M. Ruch, and Lewis M. Terman, 1940, World Book Company.

Reading Tests

1. <u>Iowa Silent Reading Tests</u>--Form AM, advanced test, new edition, (grades 9-13), H. A. Greene, A. M. Jorgensen, and V. H. Kelley, 1943, World Book Company.

2. <u>Minnesota Reading Examination for College</u> <u>Students</u>--Form A, (grades 9-16), Melvin E. Haggerty and Alvin C. Eurich, 1930, University of Minnesota Press.

3. <u>Reading Comprehension: Cooperative English</u> <u>Test</u>--Test C2, (grades 13-16), Frederick B. Davis and Harold V. King, 1942, Cooperative Test Service.

4. <u>SRA Reading Record</u>, (grades 7-12), Guy T. Buswell, 1947, Science Research Associates.

Application of Readability Formulas

The application of the formulas used in this study are discussed in the remainder of this chapter. This discussion centers on the weakness of the formulas when used for evaluating reading difficulty of standardized tests.

When the Dale-Chall and Flesch formulas were applied to the twenty-seven tests, each word was counted; there was no sampling attempted. Each statement or test

item on each of the tests that requires reading by the
person taking the test was included in the analysis of
that test. The instructions were not included in the
analyses of the tests. Instructions in many cases are read
by the examiner and explanation is given for any direction
not clearly understood by the person taking the test.
The understanding of the reading passages of the instructions
or directions for tests is primarily a function of the
examiner. However, the reading level of the instructions
for each test was to be determined later by the method
devised in this paper for evaluating readability.

Samples from each of the twenty-seven tests
were taken for analyses by the remaining three formulas.
Directions for sampling, as given by the authors of the
formulas, were followed as nearly as possible. Inter-
pretation as to the applicability was necessary in certain
instances. Lorge suggests taking three samples of one
hundred words each from long articles to be appraised,
one sample each at the beginning, middle, and end of the
article. Ten per cent of the reading material was
considered a fair sample by Yoakam; this amounted to
approximately three hundred words in the tests analyzed.
Lewerenz suggested taking a thousand word sample, so many
words from each page.

However, it seemed more appropriate to work with
the same samples from each test in order to get a fair

appraisal from each formula. Three samples of one
hundred words each were selected, one each near the
beginning, middle, and end of a test. Samples were
taken from representative areas as shown by the word
count and difficult words from the application of the
Dale-Chall formula.

This sampling was done because of the time
consuming element required by the present formulas.
Counting the total number of words and counting the
number of sentences was necessary to work the Dale-Chall,
Flesch, and Lorge formulas. Listing of each word in a
sample was required by the Lewerenz and Yoakam formulas.
The Flesch formula counts the total number of syllables
as well as the total number of words and the number of
sentences. The number of prepositional phrases are
counted in the Lorge formula. Three of the formulas
use a word list to which it is necessary to refer. The
Dale-Chall formula uses the Dale list of 3,000 familiar
words known by eighty per cent of fourth graders. The
Lorge formula uses a word list composed of the 769 words
that are common to both Thorndike's list of one thousand
words that are most frequently used and to the list of
one thousand words that are most commonly used by children
entering the first grade. The Yoakam formula refers to the
Thorndike Teachers Word List of 20,000 Words.

The time factor in working each formula is quite
high as seen in Table VI. Fourteen hours were necessary
to determine the grade placement of the Cooperative
General Culture Test by the Dale-Chall formula. The time
element became progressively less with the working of each
additional formula, except for the Yoakam formula. Some
of the elements in the calculations, such as total words,
number of sentences, and listing of difficult words, were
used by more than one formula; when these elements had
once been determined it saved time in working the other
formulas. Thus, the actual time required for each formula
worked separately would have been much longer than Table VI
shows.

A method was sought in this study to shorten the
time element necessary for finding the reading difficulty
of standardized tests. It was thought that the elimination
of the less important factors used by the five formulas
studied would greatly shorten the time required for finding
the grade level scores from the reading difficulty of
standardized tests without lessening the value of this
score.

TABLE VI

TIME IN HOURS REQUIRED TO APPLY FORMULAS

Test	Dale-Chall	Flesch*	Lorge*	Yoakam*	Lewerenz*
Adjustment Inventory	3.2	1.5	1.2	1.3	.7
California Test of Personality	4.4	2.3	1.8	2.1	1.0
Guilford-Zimmerman	5.0	3.7	2.4	3.2	1.1
MMPI	7.2	6.3	3.1	5.1	1.8
Minn. Personality	3.1	2.0	1.0	1.8	.8
Mooney	2.4	1.8	2.0	1.7	.9
Bernreuter	3.0	1.5	1.0	1.2	.8
School Inventory	2.7	1.7	1.0	1.0	.7
Study of Value	3.4	1.5	1.0	1.2	.7
ACE	3.0	1.3	1.0	1.0	.8
AGCT	3.2	2.5	1.2	2.0	.9
CTMM	3.1	2.0	1.0	1.7	.8
Henmon-Nelson	3.0	1.5	1.0	1.1	.8
Ohio State Psychol.	5.2	3.5	2.0	2.5	1.0
Otis Employment	2.2	1.5	1.0	1.1	.6
Otis Quick-Scoring	2.2	1.7	1.2	1.2	.8
Cleeton	4.1	2.5	1.4	2.2	.9
Kuder BB	3.0	3.0	1.5	2.5	1.0
Kuder CM	4.0	3.0	1.5	2.3	1.0
Lee-Thorpe	4.2	3.0	1.5	2.4	1.0
Strong	3.0	1.5	1.0	1.1	.8
Coop. Gen. Culture	14.0	10.0	3.0	4.0	1.5
Stanford Achieve.	8.0	5.2	2.5	3.4	1.2
Iowa Silent Reading	9.0	6.0	3.0	4.2	1.4
Minnesota Reading	5.0	4.2	2.0	2.0	.9
SRA Reading	5.0	2.7	1.5	1.8	1.0
Totals	143.6	82.9	44.8	53.8	26.1
Mean	5.3	3.1	1.7	2.2	1.0

* Several formulas contain some identical elements; the last formulas took less time because some of these calculations had been performed previously.

In order to apply the five formulas to the twenty-
seven tests many subjective decisions were necessary. The
Dale-Chall formula lists twenty-one rules that must be
followed in order to count the difficult words. The Lorge
formula has thirty special rules to follow in making the
count of difficult words and prepositional phrases. The
Yoakam formula lists only six rules, leaving much to the
imagination of the grader. The remaining two formulas
have no special rules to follow since they do not employ
the same meaning of a difficult word as do the formulas
that use a word list. In order to apply the special rules
one must decide when to count hyphenated words,
prepositional phrases, proper nouns, adjectives ending in
y, verbs changed to adverbs, etc.

The elimination of these factors of subjectivity
was sought in the present study in order to derive a more
objective method.

Preliminary sampling of the tests showed wide
variation among the five formulas as to the grade place-
ment of the reading difficulty of the tests. A comparative
study was made of the grade placement scores calculated
from the five formulas in order to determine whether or
not agreement could be found between any of the five
formulas. The scores derived from these formulas were
scrutinized carefully, as well as the formulas themselves,
in order to identify the weaknesses of the present read-

ability measuring instruments. Rank order correlations were used to show the agreement, not only between the formulas, but also between each formula and the mean of all five formulas for each standardized test.

The large variance found between the grade placement scores for a single test by the formulas made it clear that a more stable measure of readability was needed for tests. It was hoped that by using the mean grade score for each test as the criterion of difficulty, a method could be evolved that would overcome the variation among the formulas now in use. The mean grade placement score was found for each test by taking the grade placement for each of the five formulas and averaging them.

The proposed method to be developed in this study for determining the grade level scores of standardized tests was validated by having several people apply the technique to a sample of the tests used herein. It was proposed to see if the same approximate grade level scores could be obtained by others working independently from the writer, and to see if the technique could be applied with ease and with little consumption of time. It was further proposed to see if people unfamiliar with readability techniques, as well as those working in the field, could apply the newly developed method and obtain comparable results with ease and understanding.

Results of the above applications will be given
in the following chapter, followed by the development of a
method for judging the readability of standardized tests.
It is hoped that a technique can be developed that will
lend itself to greater ease of application, be more
objective and give more stable reading difficulty scores
for standardiaed testing instruments than do any of the
present formulas.

Summary

The five readability formulas were applied to
twenty-seven standardized tests commonly used in counseling.
Each formula gave a grade placement for each test. The
mean grade placement for each of the tests was taken as the
criterion of difficulty for that test. The vocabulary
difficulty was determined for these tests by taking the
number of words above the most frequently used four
thousand words and giving them a weight according to the
Thorndike Junior Century Dictionary. These weights were
totaled for each test and divided by the number of words
in the sample; this gave the index of vocabulary difficulty.
The indices falling within a mean grade placement were
taken as the limits for that grade.

CHAPTER V

FINDINGS AND DEVELOPMENT OF THE TECHNIQUE
FOR USE WITH STANDARDIZED TESTS

Findings

The comparative grade placement scores, as derived from the five formulas used in this study are shown in Table VII. The mean of the grade placement scores for the five formulas is given for each of the twenty-five tests as well as the mean for the four formulas which show closest agreement. The Lewerenz formula was dropped from the computation of means in order to see if there would be closer agreement between the remaining formulas and the mean grade level scores. However, after doing this, little difference was noted in the placement of the tests in rank order of difficulty.

The Yoakam formula placed the twenty-seven tests that were analyzed at higher grade levels than did any of the other formulas. The mean Yoakam grade placement score for all tests analyzed was 11.31. The Flesch formula rated the tests in the next to highest grade placement with a mean grade level of 10.76. A mean grade placement of 9.04 was derived from the Lewerenz formula. The least difficult of all grade placement scores for the twenty-seven tests were obtained from the Lorge and Dale-Chall formulas. They placed the mean test reading difficulty at 7.27 and 7.37, respectively.

TABLE VII

COMPARATIVE GRADE PLACEMENT OF SELECTED STANDARDIZED TESTS

Test	Average of Five Formulas	Average of Four Formulas	Dale-Chall	Flesch	Lorge	Lewerenz	Yoakam
MMPI	5.42	5.22	5.54	6.11	4.44	6.22	4.77
School Inventory	5.53	5.04	5.55	6.32	5.00	7.60	3.29
Calif. Test Pers.	6.20	6.32	6.18	7.18	5.34	5.72	6.57
AGCT	6.92	6.66	5.92	11.00	5.70	7.98	4.00
Guilford-Zimmerman	6.92	6.59	5.97	7.54	5.55	8.24	7.30
Otis Q-S	6.98	6.68	6.12	7.05	5.84	8.18	7.70
Adjustment Inv.	7.65	7.40	6.40	9.14	6.05	7.68	8.00
Minn. Pers. Scale	7.82	8.14	6.50	9.12	6.23	6.54	10.80
Mooney	8.12	7.85	6.10	8.30	6.03	8.68	10.95
Bernreuter	8.20	8.43	6.76	8.36	6.69	7.30	11.90
CTMM	8.30	8.37	7.18	10.81	8.84	8.03	6.66
Stanford Ach.	8.66	9.23	7.02	8.40	7.00	6.37	14.50
Kuder CM	8.74	8.98	7.25	8.53	7.66	7.82	12.46
Otis Employ.	8.81	8.71	6.33	7.92	6.29	9.30	14.28
Henmon-Nelson	9.12	8.64	6.64	9.20	6.06	11.27	12.64
Iowa Silent	9.29	9.33	8.01	11.40	7.86	9.11	10.06
Lee-Thorpe	9.45	9.87	8.03	10.02	7.86	7.78	13.56
Kuder BB	9.45	9.70	7.57	9.15	7.61	8.48	14.46
SRA Reading	10.50	10.63	8.33	13.19	8.45	9.92	12.56
Cleeton	10.87	11.53	8.36	14.38	7.36	8.26	16.00*
Strong	11.35	11.18	8.89	15.80	6.74	12.06	13.28
Coop. Reading	11.42	11.67	8.72	14.03	9.92	10.44	14.00
Minn. Reading	11.45	11.90	8.96	13.19	9.43	9.65	16.00*
Ohio State Psy.	11.53	12.12	10.69	16.50	9.57	9.08	11.80
ACE	11.70	12.26	8.50	16.10	8.45	9.44	16.00*
Coop. Gen. Cul.	12.20	12.68	8.52	15.56	10.65	10.25	16.00*
Study of Values	12.70	12.71	9.11	16.10	9.61	12.70	16.00*

* Estimate of the grade; the formulas did not indicate grades at these levels.

In addition to having the lowest reading difficulty, the greatest ranges of grade placement scores were also secured with the Yoakam and Flesch formulas. The Yoakam method gave a range of 12.71 grades, ranging from 3.29 to 16.00 grade levels of placement. The Flesch formula rated the tests from 6.11 to 16.50 grade levels, a range of 10.39 grades. The other three formulas vary little in the placement of the tests into grade levels of reading difficulty; the Lewerenz formula from 5.62 to 12.70, a range of 6.98 grades; the Lorge formula from 4.44 to 10.65, a range of 6.21 grades, and the Dale-Chall formula was the least sensitive of all, when used on this material, ranging from 5.54 to 10.69, a range of only 5.15 grades.

Large differences were also noted between the grade placement scores calculated by the various formulas for some of these tests. These differences are evident in Table VII. The difference in reading difficulty placement of the Stanford Achievement Battery, for example, was as much as 8.13 grades. The Yoakam formula rated this battery of tests in the 14.50 grade level, while the Lewerenz formula placed it in the 6.37 grade level. The large difference in the grade placement of the Stanford Achievement Battery was caused by the methods used in the two formulas rating it. The Yoakam formula rated all reading material on a higher reading level than any of the other formulas and in this particular standardized test few words

are used beginning with letters which the Lewerenz
formula rated as difficult words. The Minnesota Multi-
phasic Personality Inventory showed the smallest variance
as to grade placement scores. The highest grade level
given to this test was 6.22 by the Lewerenz formula, the
Lorge formula placed it in the 4.44 grade level. The
difference between the Lorge and Lewerenz grade placement
scores for the Minnesota Multiphasic Personality Inventory
was 1.78. However, all formulas showed little variance in
the lower grade levels.

Each formula grade placement score correlated
significantly with the mean of the five combined grade
placement scores as shown in Table IX. These high corre-
lations were partially expected, since each formula is
included in the combined grade placement scores tending to
spuriously increase the relationship. The correlations
between the various formulas were not as high as the
correlation with the mean. However, they, too, were
statistically significant.

The most sensitive formulas, those having the
greater range of grade placement scores, seemed to be the
Yoakam and the Flesch. Both of these formulas used word
difficulty as an element in grade placement of reading
difficulty; the Flesch formula counted the number of
syllables per one hundred words, while the Yoakam formula
weighted all words above the most common five thousand as
difficult.

The Lorge and Dale-Chall formulas were somewhat restricted in their range by the use of comparatively short word lists. Both of these formulas also use the factor of average sentence length in calculating the grade placement scores. The smallest range in grade placement was given by the Dale-Chall formula which used only difficult words as its other variable. Difficult words were in this case those words above the most common three thousand. As stated previously, a difficult word to the Lorge formula was any word not found in the list of 769 words common to both Thorndike's most frequently used thousand words and to the thousand words most commonly used by children entering the first grade.

Three formulas that used the average sentence length and difficult words as factors of difficulty—Dale-Chall, Flesch, Lorge—gave rank difference correlations between each other greater than .80, as shown in Table VIII. The remaining formulas—Lewerenz and Yoakam—correlated somewhat lower with the other formulas. However, all correlations were significant at greater than the .01 level as shown by Olds' Tables.[1]

---

[1] E. G. Olds. "Distribution of Sums of Squares of Rank Differences for Small Numbers." _Annals of Mathematical Statistics_, 9 (March 1938) 133-149.

CHAPTER VIII

RANK DIFFERENCE CORRELATIONS BETWEEN EACH OF
THE FIVE FORMULAS

	Dale-Chall	Flesch	Lorge	Lewerenz	Yoakam
Dale-Chall		.91	.90	.65	.75
Flesch	.91		.81	.66	.66
Lorge	.90	.81		.60	.69
Lewerenz	.65	.66	.60		.59
Yoakam	.75	.66	.69	.59	

TABLE IX

RANK DIFFERENCE CORRELATIONS BETWEEN THE MEAN GRADE LEVEL
SCORES FOR THE FIVE FORMULAS AND THE GRADE LEVEL
SCORES FOR EACH FORMULA

Formula	$\sum d^2$	ρ	p
Dale-Chall	174	.95	$< .01$
Flesch	339	.90	$< .01$
Lorge	357	.89	$< .01$
Lewerenz	737	.77	$< .01$
Yoakam	349	.89	$< .01$

The rank difference correlations between the mean grade level score for each test (Table VII) and the grade level score derived from each formula is shown in Table IX. The Dale-Chall formula grade level scores correlated .95 with the mean score from the five formulas. The lowest correlation was between the mean grade level score and the Lewerenz grade level placement score, which was .77.

It was thought that the .77 correlation between the Lewerenz grade level score and the mean scores for the five formulas might indicate that this formula was having undue adverse effect on the correlated scores. The Lewerenz formula grade level scores were dropped from the mean grade level scores and rank difference correlations were re-computed. There seemed to be little difference in these correlations versus the original ones. By omitting the Lewerenz formula from the means, two of the correlations increased by approximately .02, while two of them decreased by approximately the same figure. These comparisons are taken from Tables IX and X.

The Development of a New Technique

Since the primary purpose of this study was to develop a shorter, more objective, and more stable method of measuring the readability of standardized tests, various techniques were considered for this purpose. Sentence

length and prepositional phrases were not practical for
use as factors in determining the readability of
standardized tests because many of them have sections
composed of only lists of words instead of complete
sentences. Adjectives, adverbs, and words beginning with
certain letters all seemed to fall under the heading of
difficult words. Because of the special construction of
many standardized tests the method developed in this study
was done in the following steps:

1. Three samples of one hundred words each were
taken from the tests to be analyzed. The samples were
selected in each test at the beginning, middle, and end.
The only requirements for the samples were that they con-
sist of an even hundred words, that each sample begin with
the first word of an item, and the vocabulary tests be
omitted from the samples. The vocabulary sections of the
tests were more difficult than the other printed matter in
the tests. The vocabulary sections were omitted because
they were composed of words which are uncommon to the
person of average intelligence taking these tests and
therefore would introduce too much bias. The vocabulary
tests would be of little use if they were understood by
every one for whom the test was constructed. It seemed
only fair to omit the vocabulary sections in order to get
the average reading difficulty of the standardized tests.
It seemed easiest to begin with the first word of the first

item of a test and count the first hundred word sample,
the middle sample was selected as near the mid-point of
the test as possible, and the third sample was taken by
counting backwards from the last word of the test items
until one hundred words were counted. Should the one
hundred words end within the item, proceed counting back-
wards until the first word of an item is reached, then in
order to get exactly the one hundred words omit the number
of words over one hundred at the end of the sample.

2. Each word that appeared difficult to the
grader was written on a sheet of paper. These words were
then found in the Thorndike Junior Century Dictionary (1942).
The number following the definition in this dictionary is
the weight for that word. These numbers range from one to
twenty, representing the first twenty successive thousands
of words most commonly used in the English language. Only
words above the most frequently used four thousand words
were given a weight. Any word having a weight of four or
above was considered a difficult word and its weight was
listed. Words used more than once in the samples were given
their weight each time they were used.

3. The weights for the three samples were
totaled and divided by the number of words in the samples,
three hundred in this case. This gave the index of
vocabulary difficulty for the standardized tests. This

TABLE X

RANK DIFFERENCE CORRELATIONS BETWEEN THE MEAN GRADE LEVEL
SCORES FOR FOUR FORMULAS AND THE GRADE LEVEL SCORES
FOR EACH OF THE FOUR FORMULAS

Formula	$\sum d^2$	ρ	p
Dale-Chall	128	.97	< .01
Flesch	284	.89	< .01
Lorge	308	.91	< .01
Yoakam	430	.87	< .01

TABLE XI

GRADE LEVEL OF READING DIFFICULTY AS DETERMINED BY
THE INDEX OF VOCABULARY DIFFICULTY

Index of Vocabulary Difficulty	Grade Level
1.4510 and above	College
1.2510 - 1.4509	12th grade
1.0510 - 1.2509	11th grade
.8510 - 1.0509	10th grade
.6510 - .8509	9th grade
.4510 - .6509	8th grade
.2510 - .4509	7th grade
.0510 - .2509	6th grade
.0509 and below	5th grade

index of vocabulary difficulty was for the words above the most commonly used four thousand words.

4. The standardized tests studied were placed in rank order as determined by the mean grade level scores of the five formulas. These tests were set off into grade groups. All tests falling within one half grade level above or below the grade were considered with that grade group. For example, grade level scores of 7.50 and 8.49 would be considered of the eight grade reading difficulty. The largest and smallest indices of vocabulary difficulty falling within any one grade group were considered the limits for the grade. Table XI gives the indices of vocabulary difficulty, setting the limits for the various grade levels.

5. Using these indices of vocabulary difficulty, obtained from the above three steps, the twenty-seven tests were given grade levels of reading difficulty. To obtain grade level scores to the nearest hundredth requires interpolation by each scorer of the figures in Table XI. These grade level placement scores are given in Table XII. Included in this table also for comparative purposes are the mean grade placement scores for the five and four formulas.

6. The reading difficulty was also figured for the instructions for each of the tests analyzed. The samples in some cases included all directions to the tests when they consisted of less than three hundred words; other

samplings followed the procedure outlined above; that is,
taking one hundred word samples at three points throughout
the instructions. These scores are included in Table XII.

The working time for the various analyses of the
tests and instructions separately required on the average
of about thirty minutes by the method developed in this
study compared to approximately 2.6 hours for the other
five formulas. Thus, a considerable amount of time was
saved. There is little room for decisions to be made by the
scorer since the words are weighted in accordance with an
accepted word list. If a variant of a word or a hyphenated
word does not appear in the Thorndike Junior Century
Dictionary are given weights.

The rank order correlations between the grade level
scores, derived from the method described above, and the
mean grade level scores obtained from the five studied
formulas was .94. When the Lewerenz formula was dropped
from the means, the correlation increased to only .96. Both
of these correlations can be considered almost identical
with those obtained between the Dale-Chall formula and the
mean of the grade level scores of all five formulas. The
Dale-Chall formula had the highest correlation with the
mean of the grade level scores of the five formulas, .95.

In order to check the rank order correlations,
Table XIII was constructed where the correlation coefficients
were computed by means of the ratio of the estimated true

TABLE XII

GRADE LEVEL OF READING DIFFICULTY OF TWENTY-SEVEN SELECTED STANDARDIZED
TESTS AND INSTRUCTIONS AS DETERMINED BY THE METHOD DEVELOPED
IN THIS STUDY COMPARED TO MEAN GRADE LEVELS FOR FIVE AND
FOR FOUR OTHER FORMULAS

Tests	Index of Voc. Difficulty		Grade Levels			
	Items	Instruc-tions	Items	Instruc-tions	Average of Five Formulas	Average of Four Formulas
School Inventory	.00	.29	5.00	7.20	5.42	5.22
AGCT	.10	.08	6.24	6.12	6.92	6.66
MMPI	.14	.31	6.45	7.29	5.42	5.22
Calif. Test of Pers.	.27	.11	7.11	6.32	6.20	6.32
Guilford-Zimmerman	.34	.51	7.44	8.31	6.69	6.59
Otis Quick-Scoring	.38	.12	7.64	6.37	6.98	6.68
Adjustment Inventory	.39	.06	7.75	6.05	7.65	7.40
Minn. Pers. Scale	.61	.48	8.81	8.15	7.82	8.14
Mooney	.63	.29	8.91	7.19	8.12	7.85
OTMM	.67	.11	9.09	6.32	8.30	8.43
Otis Employment	.67	.30	9.11	6.37	8.81	8.71
Bernreuter	.67	.26	9.11	7.04	8.20	8.43
Iowa Silent Reading	.71	.19	9.27	6.69	9.29	9.33
Kuder CM	.78	.52	9.65	8.33	8.74	8.98
SRA Reading	.81	.21	9.80	6.82	10.50	10.63
Hemmon-Nelson	.82	.00*	9.84	5.00	9.12	8.64
Strong Inventory	.88	.55	10.16	8.50	11.35	11.18
Lee-Thorpe	.90	.38	10.26	7.63	9.45	9.87
Coop. Reading	.94	.40	10.42	7.76	11.42	11.67
Stanford Achievement	.98	.18	10.63	6.65	8.66	9.23
Kuder BB	.98	.43	10.65	7.89	9.45	9.70
Ohio State Psychol.	1.15	.30	11.49	7.26	11.53	12.12
Minn. Reading	1.23	.68	11.89	9.15	11.45	11.90
Study of Values	1.32	.77	12.37	9.62	12.70	12.71
Cleeton	1.35	.23	12.51	6.91	10.87	11.53
ACE	1.37	.37	12.66	7.58	11.70	12.26
Coop. Gen. Culture	2.15	.36	Coll.	7.54	12.20	12.68

* There are no words above the first four thousand.

variance to the observed variance. These correlations are shown in Table XIV. In these computations variance error was computed as one-half the variance of the distribution of differences in grade placement between the average of the five formulas, four formulas, the method in this study, and each of the five formulas. The observed variance was taken to be the average of the variance of the five formulas, four formulas, the method developed in this study, and each of the formulas studied.[1] The grade placement scores were rounded to the nearest even grade for these computations.

The rank order correlations (Tables IX and X) vary only slightly from those obtained by the ratio of the estimated true variance to the observed variance (Table XIV). In both correlation techniques the Dale-Chall formula received the highest correlation of the five formulas. The

---

1 The formulas used for these computations were:

$$V_d = i^2\left[\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2\right]$$

Where V_d is variance of difference between the grade placement scores for the means of the formulas and each formula; i is the interval of 1 grade placement score; and x is the deviation per grade placement.

$$r = 1 - \frac{V_e}{V_o}.$$

Where r is the correlation between the grade placement scores for the means of the formulas and each formula; $V_e = V_d/2$, and V_o is average variance of the means of the formulas and each formula.

rank order correlations range from .77 between the Lewerenz formula scores and the mean scores of the four formulas to .97 between the Dale-Chall formula scores and the mean scores of the five formulas. The ratio of the estimated true variance to the observed variance gave correlations ranging from .82 between the Lorge and the mean of the four formulas to .90 between the Dale-Chall formula and the mean of the five formulas.

The method developed in this study correlated higher with both the means of the five and four formulas by the ratio technique than did any of the other formulas. In the rank order correlations with the means of the formulas the Dale-Chall formula exceeded this method by .004 with the mean of the five formulas and by .011 with the mean of the four formulas.

The correlations between the grade placement scores obtained by the method developed in this study and the scores from the other formulas are much higher than the correlations between the scores from any two of the other formulas. The correlations between the newly developed method scores and the other formula scores range from .95 for the Dale-Chall to .72 for the Lewerenz, Table XIV. The other formulas range in correlation with each other from .91 between the Flesch and Dale-Chall formulas to .59 between the Yoakam and Lewerenz formulas. The method developed in this study correlated the highest with the formula that

TABLE XIII

TABLE OF SCORES, DEVIATIONS AND VARIANCES USED FOR CALCULATING
THE RATIO OF ESTIMATED TRUE VARIANCE TO THE
OBSERVED VARIANCE CORRELATION COEFFICIENTS

Formula	X	X^2	x			x^2			σ	σ^2
			5	4	Fo	5	4	Fo		
Ave. of 5	245	2337							2.05	4.20
Ave. of 4	249	2443							2.30	5.29
Forbes	255	2535	26	20		32	24		2.16	4.67
Dale-Chall	201	1545	46	52	23	94	122	29	1.34	1.80
Flesch	288	3384	49	47	57	155	137	195	3.39	11.49
Lorge	198	1526	49	53	57	112	143	179	1.66	2.76
Lewerenz	234	2100	31	39	43	61	91	125	1.63	2.66
Yoakam	308	2910	75	69	65	295	245	207	3.81	1.51

Note: The symbols used:

X Sum of the grade placement scores.

X^2 Sum of the squared grade placement scores.

x Sum of the differences between the grade placement scores.

x^2 Sum of the squared differences between the grade placement scores.

σ Standard deviation of the grade placement scores.

σ^2 Variance of the grade placement scores.

correlates the best with the other formulas and the lowest with the formula that correlates the least, at the same time showing close relationship with the mean grade placement scores of the formulas.

In order to check the reliability of the newly developed formula four scorers applied the method to nine of the twenty-seven tests. The tests scored for reading difficulty were chosen from varying levels of difficulty. The three tests rated the easiest by the mean score of the five formulas, the three falling at the mid-point of difficulty, and the three most difficult tests were selected. Two of the scorers were familiar with problems involved in reading and the other two individuals had no specialized training in reading. The instructions for figuring the readability by this method as outlined previously in this chapter were handed to the individuals with no further directions. Each scorer figured the reading difficulty for each of the nine tests. The grade placement scores varied little from scorer to scorer as seen in Table XV. The graders, following instructions above, each worked with the same first and third samples. Their second (middle) sample of one hundred words varied since it could be taken any place near the mid-point of the tests. Table XVI shows the mean grade placement scores determined by the four scorers compared to the grade placement scores obtained by the writer.

TABLE XIV

CORRELATION COEFFICIENTS FOR GRADE PLACEMENT BY READABILITY
FORMULAS VS. AVERAGE GRADE PLACEMENT BY FIVE AND FOUR
FORMULAS RESPECTIVELY USING RATIO OF ESTIMATED
TRUE VARIANCE TO OBSERVED VARIANCE

Formula	Ave. of Five	Ave. of Four	Forbes
Dale-Chall	.90	.88	.95
Flesch	.84	.88	.83
Lorge	.87	.82	.84
Lewerenz	.86	*	.72
Yoakam	.87	.87	.90
Forbes	.96	.97	

* The Lewerenz formula was dropped for this calculation.

TABLE XV

VOCABULARY DIFFICULTY INDICES AND GRADE PLACEMENT
SCORES FIGURED BY THE FOUR INDEPENDENT WORKERS

Test	Index of Difficulty				Grade Placement			
	1	2	3	4	1	2	3	4
School Inventory	.04	.05	.04	.04	5.00	5.00	5.00	5.00
AGCT	.08	.09	.08	.08	6.13	6.20	6.15	6.13
MMPI	.12	.13	.12	.13	6.35	6.40	6.36	6.40
Kuder CM	.73	.80	.82	.77	9.41	9.75	9.84	9.58
Iowa Silent	.75	.77	.85	.73	9.50	9.60	10.01	9.41
SRA Reading	.84	.81	.83	.80	9.95	9.80	9.92	9.77
Cleeton	1.17	1.37	1.18	1.45	11.62	12.58	11.67	12.98
ACE	1.32	1.34	1.37	1.45	12.35	12.45	12.60	12.98
Coop. Gen. Cul.	1.83	1.95	1.90	2.00	Coll.	Coll.	Coll.	Coll.

TABLE XVI

AVERAGE INDICES OF VOCABULARY DIFFICULTY AND GRADE
PLACEMENT SCORES BY OTHER GRADERS COMPARED
TO SCORES DERIVED FROM THIS STUDY AND
DIFFERENCE IN GRADE PLACEMENT

Test	Other Graders		The Writer		Difference in Grade Placement
	Index of Voc. Diff.	Grade Level	Index of Voc. Diff.	Grade Level	
School Inventory	.04	5.00	.00	5.00	.00
AGCT	.08	6.15	.10	6.24	.09
MMPI	.13	6.38	.14	6.45	.07
Kuder CM	.78	9.64	.78	9.65	.01
Iowa Silent	.78	9.63	.71	9.27	.26
SRA Reading	.82	9.86	.81	9.80	.06
Cleeton	1.29	12.21	1.35	12.51	.30
ACE	1.37	12.60	1.37	12.66	.06
Coop. Gen. Cul.	1.92	Coll.	2.15	Coll.	.00

The tests were placed in the same position in rank order by these scorers as was found by the writer. However, there was slight variation as to the grade levels for the tests, two tests varying as much as .26 of a grade. The various sections of the <u>Iowa Silent Reading Test</u> vary greatly as to their reading difficulty. Parts of the <u>Cleeton Vocational Interest Inventory</u> also vary widely as to the words used for occupations, activities, school subjects, and personal traits. The writer chose samples which appeared to be of average difficulty from each of the tests when figuring their reading difficulty. The Flesch and Dale-Chall formulas were figured for each page of each test, thus making it possible to choose passages of average reading difficulty. The samples chosen by the scorers were not necessarily representative of the reading difficulty of the tests. This accounts for the slight difference among the grade level scores obtained by the other scorers.

Summary

This study was undertaken specifically to find the reading level of standardized tests and to develop a new method for doing this. The method developed was as sensitive as any of the five methods studied in this paper. The grade placement scores ranged from the fifth grade to college

level of reading difficulty.   The method developed was
based entirely upon the type of reading matter found in
standardized tests.

Thus, using only the difficult words of the
reading matter in the standardized tests, the writer
could achieve as valid and reliable readability scores as
were secured when many factors were included. The time
element was reduced greatly and greater objectivity was
obtained with the elimination of most of the interpretations
or judgments formerly left to the scorers.

CHAPTER VI
SUMMARY AND CONCLUSIONS

Summary

This study was to be a comparative analysis of the five most recent techniques devised for the determination of the inherent difficulty of reading materials in order to obtain a method specifically suited for measuring the reading difficulty of standardized tests. The five formulas, Dale-Chall, Flesch, Lorge, Lewerenz, and Yoakam, were examined for the elements of reading difficulty they measured. The formulas were applied to twenty-seven selected standardized tests commonly used in counseling. A definite difference was noted in the results of the measurement of the various tests by these five formulas. There was as much as 8.13 grades difference in the reading difficulty of a single test as determined by two different formulas.

At the same time the five formulas correlated significantly with each other. The rank order correlations ranged from .91 between the Dale-Chall and Flesch formulas to .59 between the Lewerenz and Yoakam formulas.

The grade levels determined by each formula for each test were averaged in order to obtain the mean grade placement score for each test. The rank order correlations between each of the formulas and this mean grade level score

ranged from .95 for the Dale-Chall formula to .77 for the Lewerenz formula. Since the correlation between the Lewerenz formula scores and the mean grade level scores was so low, the Lewerenz formula was dropped from the averages and the correlations recomputed. When the mean grade level scores of the remaining four formulas were correlated with each formula grade placement score, there was found to be very little difference from the original correlations.

Correlations were computed by means of the ratio of the estimated true variance to the observed variance between each formula and the means of the five and four formulas. These correlations ranged from .90 between the Dale-Chall formula and the mean of the five formulas to .82 between the Lorge formula and the mean of the four formulas. However, the scores obtained from this study for the reading level of each test correlated slightly over .95 with both the means of the five and the four formulas.

The amount of knowledge about grammar needed to apply most of these five techniques studied make them extremely difficult. Also, the amount of time required to work the methods made them too laborious. More than ten hours were required to apply some formulas to a single test. The average amount of time for the working of a single formula on a single test was more than two and one-half

hours. The newly developed method in contrast requires approximately one half hour per test.

The analyses of these five techniques for judging readability were made in order to determine some method of evaluating the readability of standardized tests. A method was sought which would greatly lessen the time required to determine the readability of tests, which could be applied by laymen, and at the same time give a more objective grade level placement score for standardized tests.

Word difficulty was used as a common factor to all five formulas studied. It was also evident from the review of the literature that word difficulty was basic to the readability of all printed matter.

The following steps were taken for the development of a method specifically suited for finding the readability of printed matter in standardized tests:

1. The five methods studied were applied to each of the twenty-seven standardized tests.

2. The mean grade level score of the five formulas for each test was taken as the criterion of difficulty for the tests.

3. The vocabulary difficulty was determined for each test by finding the number of words above the most frequently used four thousand words in three samples of one hundred words each selected at the beginning, middle, and end of each test.

4. The <u>Thorndike</u> <u>Junior</u> <u>Century</u> <u>Dictionary</u> was used for finding the weights to be assigned to each word above the most commonly used four thousand. These weights follow the definitions in this dictionary. The weights range from one to twenty, but since the first four thousand were dropped, only numbers of four and above were used.

5. The total of these weights for each test was divided by the number of words in the samples, giving the index of vocabulary difficulty.

6. The indices falling within a mean grade placement, i.e., 6.50 to 7.49, were taken as the limits for that grade.

7. The grade level scores derived from this method give the reading grade level at which the printed matter in the standardized test should be read with understanding by the person taking the test.

Conclusions

The conclusions found, within the limitations of this study, were:

1. The review of the literature showed that no specific method has been developed for finding the reading difficulty of standardized tests up to the present time.

2. The five techniques for measuring the readability of printed matter that were applied to the twenty-

seven standardized tests in this study showed wide vari-
ation as to the grade placement of the reading difficulty
of these tests.

3. The methods in use for determining the
readability of reading material consume a great amount of
time for their application.

4. These methods also required much interprotation
and judgment on the part of the user; thus greatly lessening
their objectivity.

5. The peculiar make-up of the reading matter
in the standardized tests required that only the vocabulary
difficulty factor be used for determining their readability.
The use of such factors as sentence length and prepositional
phrases was not practical since many of the tests have
sections composed only of word lists.

6. The instructions to the standardized tests
were easily within the range of reading difficulty of those
for whom the tests were designed.

7. The use of short word lists for determining
difficult words tended to give too coarse a classification
of grade levels of reading. A longer list made the method
for determining the readability of standardized tests more
sensitive, spreading the grade level scores out over a
longer range.

8. The method developed in this study was based
entirely upon the reading matter found in commonly used

standardized tests. It is a technique applicable only to such reading matter.

9. The method evolved from this study is easily applied, consumes very little time, and shows high objectivity by the elimination of most of the interpretations and judgments formerly left to the scorer.

BIBLIOGRAPHY

1. Allard, J. A. "Difficulty of Poems Commonly Presented to Elementary School Pupils." *University of Pittsburgh Bulletin*, 42 (Fall 1945) 9-18.

2. Aukerman, R. C., Jr. "Differences in Reading Status of Good and Poor Eleventh Grade Students." *Journal of Educational Research*, 42 (March 1948) 498-515.

3. Baker, G. and Peatman, J. H. "Tests Used in Veterans Administration Advisement Units." *American Psychologist*, 2 (March 1947) 99-102.

4. Berger, Herman I. "The Difficulty of Third-Grade Health Readers." *The Elementary School Journal*, 47 (March 1947) 391-395.

5. Berkshire, J. R., Bugental, J. F. T., Cassens, Frank P., and Edgerton, Harold A. "Test Preference in Guidance Centers." *Occupations*, 26 (March 1948) 337-343.

6. Betts, Emmett A. "Readability: Its Application to the Elementary School." *Journal of Educational Research*, 42 (February 1949) 438-359.

7. Bond, Guy L. "Identifying the Reading Attainment and Needs of Students." *Yearbook of the National Society for the Study of Education*, (1948-2) 69-90.

8. Brayfield, Arthur H. and Reed, Patricia A. "How Readable Are Occupational Information Booklets?" *Journal of Applied Psychology*, 34 (October 1950) 325-328.

9. Brophy, D. F. and Long, L. "Veterans Administration Vocational Training Program: Processing Procedures Used by the College of the City of New York." *Psychological Bulletin*, 41 (November 1946) 109-116.

10. Brown, Robert. "Vocabularies of History and Reading Textbooks." The Principle and Supervision, *Tenth Yearbook of the Department of Elementary School Principals*, (1931) 408-411.

11. Buckingham, B. R. and Dolch, E. W. *A Combined Word List*. New York: Ginn and Company, 1936.

12. Buros, Oscar Krisen, editos. The Third Mental Measure-
      ments Yearbook. New Brunswick: Rutgers Uni-
      versity Press, 1949.

13. Burtt, Harold E. "Typography and Readability."
      Elementary English, reprint, (January-May 1949).

14. Chall, Jeanne S. "This Business of Readability."
      Educational Research Bulletin, 26 (January 1947)
      2-7.

15. Christensen, T. E. "Some Observations with Respect to
      The Kuder Preference Record." Journal of Edu-
      cational Research, 40 (October 1946) 96-107.

16. Cole, Luella. The Teachers Handbook of Technical
      Vocabulary. Bloomington, Illinois: Public School
      Publishing Co., 1940.

17. Cowing, Amy Bronna. "They Speak His Language." Journal
      of Home Economics, 37 (October 1945) 478-489.

18. Crissy, William J. E. "How Readable Are Employee
      Handbooks?" Personnel Psychology, 4 (Winter 1951)
      383-395.

19. Dale, Edgar. Bibliography of Vocabulary Studies.
      Bureau of Educational Research, Ohio State Uni-
      versity, 1949.

20. Dale, Edgar. "A Comparison of Two Word Lists." Edu-
      cational Research Bulletin, 10 (December 1931)
      484-498.

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

22. Dale, Edgar and Chall, Jeanne S. A Formula for
      Predicting Readability. Bureau of Educational
      Research, Ohio State University, 1948.

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

24. Darley, J. G. and Marquis, D. G. "Veterans' Guidance
      Centers: A Survey of Their Problems and Activities."
      Journal of Clinical Psychology, 2 (January 1936)
      109-116.

25. Dolch, E. W. "The Use of Vocabulary Lists in Predicting Readability and in Developing Reading Materials." Elementary English, 26 (March 1949) 142.

26. Edgerton, Ronald B. "How Difficult Are Children's Encyclopedias? I and II." The Elementary School Journal, 45 (March 1945) 378-385.

27. Elliott, Catherine J. "A Critical Analyses of the Objective Methods of Measuring Reading Difficulty." Pittsburgh Schools, 25 (May-June 1941) 201-209.

28. Flesch, Rudolf F. Marks of Readable Style. Teachers College Contributions to Education, No. 897. Teachers College, Columbia University, 1943.

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

30. Fries, Charles C. and Traver, A. Aileen. English Word Lists. Washington, D.C.: American Council on Education, 1940.

31. Gates, Arthur I. "Frontiers in Educational Research in Reading." Journal of Educational Research, 40 (January 1947) 381-388.

32. Gates, Arthur I. Interest and Ability in Reading. New York: The Macmillan Company, 1930.

33. Gates, Arthur I. A Reading Vocabulary for Primary Grades. New York: Bureau of Publications, Teachers College, Columbia University, 1935.

34. Gray, William S. "The Nature and Extent of the Reading Problems in American Education." Educational Records Supplement, (1938) 87-104.

35. Gray, William S. "Progress in the Study of Readability." The Elementary School Journal, 47 (May 1947) 491-499.

36. Gray, William S. Summary of Investigations Relating to Reading. Monographs No. 28. Chicago: University of Chicago Press, 1925.

37. Gray, William S. "Summary of Reading Investigations. July 1, 1949, to June 30, 1950." Journal of Educational Research, 44 (February 1951) 401-441.

38. Gray, William S. "Summary of Reading Investigations. July 1, 1950, to June 31, 1951. Journal of Educational Research, 45 (February 1952) 412-540.

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

40. Hartmann, George W. "Further Evidence on the Unexpected Large Size of Recognition Vocabularies Among College Students." Journal of Educational Psychology, 37 (October 1946) 436-439.

41. Holland, Benjamine F. "The Effect of the Length and Structure of Sentences on the Silent Reading Process." Psychological Bulletin, 20 (November 1933) 668-669.

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

43. Johnson, Ralph H. "The Problem of Veterans Reading Level in the Counseling Process." Minnesota Counselor, (1948) 3.

44. Johnson, Ralph H. and Bond, Guy L. "Reading Ease of Commonly Used Tests." Journal of Applied Psychology, 34 (October 1950) 319-324.

45. Keboch, R. D. "Variability of Word Difficulty in Five American History Textbooks." Journal of Educational Research, 40 (January 1927) 20-29.

46. Klare, George R. Evaluation of Quantative Indices of Comprehensibility in Written Communication. Unpublished Ph.D. Thesis. University of Minnesota, 1950.

47. Latimer, Edward H. A Comparative Study of Recent Techniques for Judging Readability. Unpublished Ph.D. Thesis. University of Pittsburgh, 1947.

48. Leary, Bernice E. "Difficulties in Reading Material." Reading in General Education. American Council on Education, 1940.

49. Lewerenz, Alfred S. "Measurement of the Difficulty of Reading Material." Educational Research Bulletin, (Los Angeles City Schools), 8 (March 1929) 11-16.

50. Lewerenz, Alfred S. Techniques for the Objective Evaluation of the Vocabulary Used in Printed Matter. Unpublished Ph.D. Thesis. University of Southern California, 1932.

51. Lewerenz, Alfred S. "A Vocabulary Grade Placement Formula." Journal of Experimental Education, 3 (March 1935) 236.

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

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

54. Lorge, Irving and Blau, Ralph. "Reading Comprehension of Adults." Teachers College Record, (1941) 189-198.

55. Lorge, Irving and Kruglov, Lorraine. "The Relationship Between theReadability of Pupils' Compositions and Their Measured Intelligence." Journal of Educational Research, 43 (February 1950) 467-474.

56. McCall, William A. and Crabbs, Lilah M. Standard Test Lessons in Reading. Books II, III, IV, and V. New York: Bureau of Publications, Teachers College, Columbia University, 1926.

57. McClusky, Howard Y. "Quantitative Analysis of the Difficulty of Reading Materials." Journal of Educational Research, 28 (December 1934) 276-282.

58. MacLatchy, Josephine H. and Wardwell, Frances. "Common Pre-Primer Words." Educational Research Bulletin, 27 (November 1948) 199-226.

59. Miller, Leo R. "Reading-Grade Placement of the First Twenty-Three Books Awarded the John Newberry Prize." The Elementary School Journal, 46 (March 1946) 394-399.

60. Morriss, Elizabeth C. and Holverson, Dorothy. "Idea Analysis Technique." Unpublished Manuscript, Teachers College, Columbia University, 1938.

61. Ojemann, Ralph H. "The Reading Ability of Parents and Factors Associated with Reading Difficulty of Parent-Education Materials." Research in Parent-Education II, University of Iowa Studies in Child Welfare, 7 (1934) 9-32.

62. Olds, E. G. "Distribution of Sums of Squares of Rank Differences for Small Numbers." Annals of Mathematical Statistics, 9 (March 1938) 133-149.

63. Patty, W. W. and Painter, W. I. "Improving Our Methods of Selecting High School Textbooks." Journal of Educational Research, 24 (June 1931) 23-32.

64. Rinsland, Henry D. A Basic Vocabulary of Elementary School Children. New York: Macmillian Company, 1947.

65. Roeber, Edward C. "A Comparison of Seven Interest Inventories with Respect to Word Usage." Journal of Educational Research, 42 (September 1948) 8-17.

66. Seegers, J. Conrad and Seashore, R. H. "How Large are Children's Vocabularies?" Elementary English, 26 (April 1949) 181-194.

67. Stefflre, B. "The Reading Difficulty of Interest Inventories." Occupations, 26 (November 1947) 95-96.

68. Sterner, Alice P. Radio, Motion Pictures and Reading Interest. Contribution to Education, No. 932. New York: Bureau of Publications, Teachers College, Columbia University, 1945.

69. Thorndike, Edward L. The Teachers Word Book. New York: Teachers College, Columbia University, 1921.

70. Thorndike, Edward L. A Teachers Wordbook of 20,000 Words. New York: Bureau of Publications, Teachers College, Columbia University, 1931.

71. Thorndike, Edward L. Thorndike Century Junior Dictionary, Revised Edition. New York: Scott, Foresman and Company, 1942.

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

73. Tiffin, Joseph and Walsh, Francis X. "Readability of Union-Management Agreements." _Personnel Psychology_, 4 (Winter 1951) 327-337.

74. Tinker, Miles A. and Peterson, Donald G. "Reader Preference and Typography." _Journal of Applied Psychology_, 26 (January 1942) 38-40.

75. Traxler, Arthur E. "Research in Reading in the United States." _Journal of Educational Research_, 42 (March 1949) 481-499.

76. Traxler, Arthur E. _Ten Years of Research in Reading_. Educational Records Bulletin, No. 32. Educational Records Bureau, 1941.

77. Traxler, Arthur E. and Townsend, Agatha. _Another Five Years of Research in Reading_. Educational Records Bulletin, No. 46, Educational Records Bureau, 1946.

78. Vogel, Mable and Washburne, Carleton. "An Objective Method of Determining Grade Placement of Children's Reading Material." _Elementary School Journal_, 28 (January 1928) 373-381.

79. Washburne, Carleton and Morphett (Vogel), Mabel. "Grade Placement of Children's Books." _Elementary School Journal_, 37 (January 1938) 335-364.

80. Washburne, Carleton and Vogel, Mabel. _Winnetka Graded Book List_. Chicago: American Library Association, 1926.

81. West, Michael P. _Bilingualism_. Bureau of Education, India, 1926.

82. Yoakam, Gerald A. "The Reading Difficulty School Text-books." _The Elementary English Review_, 22 (December 1945) 304-309.

83. Yoakam, Gerald A. _A Technique for Determining the Difficulty of Reading Material_. Unpublished Material. University of Pittsburgh, 1939.

84. Zeller, Dale. _Relative Importance of Factors of Interest in Reading Materials of Junior High Pupils_. Contributions to Education, No. 841. New York: Bureau of Publications, Teachers College, Columbia University, 1941.

ABSTRACT

Purpose

This study was carried out in order to determine objectively the reading difficulty of standardized tests commonly used in counseling and to develop a new method for determining the reading level of these standardized tests.

Within recent years much research has been carried on in the field of reading, and more specifically in determining the readability of printed matter. Reading material of many types has been tested for readability, but the determination of the reading level of standardized tests is one area in which little has been done.

Methods

Five of the more popular techniques for evaluating the reading difficulty of printed matter were critically analyzed. The formulas studied were the Dale-Chall, Flesch, Lorge, Lewerenz, and Yoakam. These formulas were applied to twenty-seven selected standardized tests commonly used for counseling.

The grade level scores obtained for each test from the five formulas were averaged in order to obtain a mean grade level reading difficulty score for each test. These mean scores were taken as criterion grade level scores of

reading difficulty for these selected tests.

Standardized tests are not generally constructed in the same manner as most reading matter. In many cases sections of the tests are composed only of word lists instead of complete sentences denoting a train of thought or continuity of expression. Because of this peculiar make-up of testing material, it was deemed appropriate to use the vocabulary difficulty of the standardized tests in order to obtain a fair measure of their reading difficulty.

Three samples of one hundred words each were taken from each standardized test. Any word appearing in these samples that was above the most commonly used four thousand words, according to the Thorndike Junior Century Dictionary, was given a weight. These weights were the numbers following the definitions which represented the range and frequency of use of the word defined. These numbers range from one to twenty. Only words above the most commonly used four thousand were given weight, thus the weights used for this study ranged from four to twenty. The weights were totaled and divided by the number of words in the sample in order to determine the test's index of vocabulary difficulty.

The indices of vocabulary difficulty falling within the limits of any mean grade level, as determined by the average of the five formulas, were taken as the limits of their respective grade levels.

In order to determine the usefulness of the
newly developed method, four people were asked to score
nine of the selected tests from three different levels of
difficulty; the three tests having the easiest reading
difficulty, the three most difficult, and the three that
fell at the mid-point of difficulty were chosen for this
grading. Their grade level placement scores for these
tests were in close agreement with those obtained by the
writer. The tests were placed in their same relative order
of difficulty and only two of them were given reading grade
levels as much as a quarter of a grade different from that
found by the writer.

Results

This study in readability resulted in a new method
for determining the readability of standardized tests. The
method developed was based solely on the printed matter
found in testing instruments and is to be used specifically
to find the readability of such material.

This method gave reading grade levels for the
standardized tests that approximated the mean grade level
scores of the five formulas studied. A large amount of
time was saved by the method developed because it cut the
time to one-fifth of that required by the methods previously
developed.

Objectivity of measurement of the readability of these tests was greatly increased. Most of the interpretations and judgments left to the scorers by the other methods were eliminated. To find the reading difficulty of any standardized test the scorer needs only to use the Thorndike Junior Century Dictionary for the weights of words above the most commonly used four thousand words and follow the steps listed below:

1. Choose three samples of one hundred words; one sample being taken each from the beginning, middle, and end of the test items.

2. Write the words that appear difficult on a piece of paper and look the weights up for these words. When in doubt as to the difficulty of a word, find its weight. Do not be fooled by adverbial forms.

3. Total the weights that are four and over for the three samples.

4. Divide the sum of the weights by the number of words in the sample. This is the index of vocabulary difficulty.

5. Refer to Table XI, Grade Level of Reading Difficulty as Determined by the Index of Vocabulary Difficulty, for the grade placement of the material sampled. Interpolate to the nearest hundredth of a grade.

Conclusions

The general conclusions evident from this study were:

1. Readability of standardized tests was a neglected area in reading research.

2. Formulas and methods in current use gave wide variation of reading grade level scores for printed matter in standardized tests.

3. An index of vocabulary difficulty was the sole factor usable for determining the readability of such reading material because of the peculiar construction of testing instruments.

4. Methods that were in use for determining the readability lacked a high degree of objectivity, leaving a large amount of judgment and interpretation to the user.

5. A large amount of time was also required for the application of these previously developed methods.

6. The newly developed method for finding the difficulty of the reading in standardized tests overcomes these objections to a large extent. Greater objectivity was achieved by the elimination of interpretations and judgments left to the users. The time element for the application of the developed method consumed but one-fifth of the time necessary to apply the formerly developed methods. This method was developed specifically from the

type of reading found in testing instruments and gave a
measure of readability for this type of reading matter.