Readership and readability.

Palermo, Michael Joseph
Boston University

http://hdl.handle.net/2144/9099
Boston University
READERSHIP AND READABILITY

A Thesis
Presented to
the Faculty of the Department of Journalism
Boston University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

by
Michael Joseph Palermo
June 1954
Approved
by
First reader. David M. White
Professor of Journalism

Second reader. Nathan Mace
Professor of Research
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. The Problem</td>
<td>1</td>
</tr>
<tr>
<td>Definition of Concept Readability</td>
<td>4</td>
</tr>
<tr>
<td>Definition of Concept Readership</td>
<td>5</td>
</tr>
<tr>
<td>II. Readability formulas</td>
<td>6</td>
</tr>
<tr>
<td>III. Styles of Type Faces as Factors of Readability</td>
<td>63</td>
</tr>
<tr>
<td>IV. Headlines--Their Functions and Purposes</td>
<td>72</td>
</tr>
<tr>
<td>V. Readership: Its Components and Measurements</td>
<td>87</td>
</tr>
<tr>
<td>VI. Reader-Interest Survey Techniques</td>
<td>126</td>
</tr>
<tr>
<td>VII. Summary and Conclusions</td>
<td>137</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>149</td>
</tr>
</tbody>
</table>
CHAPTER I

The Problem. This is primarily an evaluation of the literature on readership and readability in order to determine the validity and reliability of claims of findings by researchers who feel they are scientifically oriented in their methodology.

The original intention of the writer was to evaluate the literature within the area of psychology, that is, what has psychology done for journalism. The basic objective of this science, as any other, is to observe, predict, and when possible, to control the variables existent in any situation or society. After perusing the Psychological Abstracts for the last twenty-five years, however, the original objective had to be modified in light of the paucity of material listed in these abstracts. It was found that most research of a quasi-psychological nature pertaining to journalism and its sub-divisions was located in journalism and public opinion quarterlies. These researchers, in any event, must employ psychological principles and scientific methodology to obtain validity for their findings.

The role of journalism, or more specifically, the newspaper, is to be an unbiased, impartial recorder. Generally speaking, the modern newspaper fulfills such a
role. But it is a truism that the newspaper does influence society, and society influences the newspaper. The point of greatest controversy lies in the degree to which one influences the other. The modern paper is what it is because the organization of modern society is what it is. Changes in the newspaper can be understood only in light of changes in the social environment of which it is a part.

The newspaper establishment is an economic enterprise. Daily and weekly newspapers are businesses which rely for financial success on circulation and advertising. As changes in the social environment take place, the newspaper must change along with it, or perish. Publishers have utilized many generalities of human behavior, most of which have been empirically determined. Journalism researchers have conducted a considerable number of studies to ascertain what improvements could be made upon the publisher's empirically determined procedures. Researchers have sought to find what are the most effective changes to make (readability) that will appeal to the wishes, wants, and needs of the reader (readership).

Few psychologists (Tinker, Paterson, the Allports) have conducted newspaper researches. Some of the experiments

---

1 Malcolm W. Willey, "The Influence of Social Change on Newspaper Style," Sociology and Social Research, 13:30-37, 1928
which have been done on readability—reading as a visual task, the eyelid reflex as a criterion of ocular fatigue, the readability of certain type sizes and forms in sight saving classes, studies of visual fatigue—are beyond the practicability for adoption by the newspaper.

Social psychologists, when they do study newspapers, are mainly concerned with the influence of social change on the newspaper, or the newspaper as a vehicle of propaganda for certain classes or groups. Occasionally, they quote an experiment in a class room situation showing how a slanted editorial affected the attitude of the students toward some individual. The primary purpose, however, is to indicate the role of newspaper as a medium of propaganda.

Under the title of researcher are men from various fields—commercial research concerns, public opinion specialists, journalists, journalism instructors, public relations instructors. In the earlier phases of the history of research in readership and readability, most workers were not scientifically oriented, and few significant contributions were made. Findings of one researcher frequently conflicted with the findings of another. Recently, most researchers


have been using sophisticated statistical tools and have been following a somewhat rigorous experimental procedure.

Readability

The concept of readability covers three major areas of research: styles of type faces; ease of understanding an article (Flesch and other readability formulas), and headlines.

**Styles of type faces:** Most of the research done in this field has been by Tinker and Paterson. They have experimented with illumination intensities for reading newspaper type, differences among newspaper body types in readability, influence of leading upon readability of newspaper type. Also included in this section is an article by Barnhart and Jones concerning read reaction to nine-column newspaper and page shrinkage.

**Readability formulas.**

Formulas are applied to reading matter to determine the level of difficulty for the reader. One most widely used, and most controversial, was devised by Flesch, revised by Flesch and others. Some writers have pointed out that an easy-to-read score by the Flesch formula does not always mean that the article is easily comprehended, a failing which Flesch, himself, admits.

**Headlines**

Research on headlines covers such areas as its function, family types, size, effects of headlines in creating a "set."
Readership

The concept of readership is the interests of the readers in various newspaper articles, such as pictorial content, general news and information, comment, sports news, practical guidance, fiction, and advertising. Reading habits are also evaluated.

The nature of readership is such that it does not lend itself to experimental procedure to the same degree as readability studies. Researchers have to utilize mailed questionnaires and polling techniques. Therefore, some articles discussing the validity of polling procedures, the interview vs. the mailed questionnaire are included.

Definition of terms.

coefficient of correlation - the amount of similarity, in direction and degree, of variations in corresponding pairs (r) (As used by Nafziger, Weigle, Flesch, Hayes, Jenkins, Walker, Jones, Bush, and Teilhet)

validity - a study is valid to the degree that it is measuring what it is supposed to measure.

reliability - no test or research can have validity unless it measures accurately. The accuracy of measurement is expressed in the reliability coefficient.

factor analysis - By observation intercorrelations of reading in the various categories, we can learn much about the basic organization of such reading. A factor is a generalization of a group of variables which are highly interrelated in the population under study. A certain amount of the variation of each variable in the group is explained by the variation of a single common factor. Factor analysis is simply a statistical method for defining a number of relatively specific variables in more general, more economical terms.

analysis of variance - the isolating from comparable groups of data the variations traceable to specified sources. (p. 79)
CHAPTER II
READABILITY FORMULAS

What is readability as measured by a formula? Short sentences, short words, condensation, human interest, conciseness, lucidity, comprehensibility are but a few of the many definitions presented in recent years. With the advent of readability formulas numerous articles have appeared hypothesizing, theorizing, experimenting, upon formulas and devices which their authors thought would tap that concept called readability. The most popular of these has been the Flesch ease of reading formula.

Most comparisons of measuring rods designed to measure readability support directly or inferentially the theory that shorter words and sentences make for increased comprehensibility. This does not mean, of course, that all short words are "easy" or that the comprehensibility of sentences depends altogether on their length. It simply means that a short word or short sentence is more likely to be comprehended readily than a long one.¹

Three conclusions, at least, can be drawn about the readability formulas in current use:

1. It is a mistake to expect the scores produced

---
by these formulas to correlate with actual readership or to provide a valid basis for predicting over-all readability. The formulas deal primarily with writing style. They deliberately omit the study of content, format, and organization, each of which is important to readability.

2. It is a mistake to consider any of the formulas as a recipe for a good writing style. They consider only a few elements of writing style; they overlook a great many others.

3. Certain mechanical characteristics of the formulas—the advantage they give to conversational type material, their inflexible allowance for changes in style—demand that they be used with discretion.²

But we must not throw out the baby along with the dirty bath water. The formulas are of considerable importance. They are steadily being improved upon by notable researchers in the field.

Views by various researchers on the subject of readability will be presented, followed by a series of investigations on one very widely used, the Flesch formula.

A few experiments have tested the way in which two or more style factors together are related to readership—such factors as sentence length, vocabulary, and so on. These experiments have shown, at least for their particular audiences, that readership gains result from increased readability.3

In the spring of 1949, a two part study by Ludwig4 designed to isolate and test the relation of specific style factors to readership was completed at the State University of Iowa in cooperation with Wallaces' Farmer. The first experiment tested the related of "hard" words to readership. A story on the subject of hog feeding was written at the level of 132 syllables per hundred words. Then the story was rewritten at the 158-syllable level. (According to the Flesch formula, the 132-syllable version was easier to read than the 158-syllable version.)

To find out how "human interest" affects readership, a second story was written—this time on the topic of landlord-tenant relations among farmers. This story, Cl, was written at the Flesch "human interest" level of 72,

3 Donald R. Murphy, "How Plain Talk Increases Readership 45% to 66%," Printers' Ink, 220:35-37, 1947, also, Charles E. Swanson, "Readability and Readership," Journalism Quarterly, 25:339-343, 1948

4 Merritt C. Ludwig, "Hard Words and Human Interest; Their Effects on Readership," Journalism Quarterly, 26:167-171, 1949
"dramatic." A rewritten version, C2, was at the 29 level, "interesting."

Since the "human interest" score depends on the number of "personal" words and sentences, C1 and C2 varied in these factors. Differences were produced by the word-substitution method—for instance, "the corn" instead of "his corn." A change in the number of "personal" words automatically changed the number of "personal" sentences, since "personal" words are one factor which make sentences "personal." 5

Ludwig's use of the interest-control stories was based on two assumptions:

1. Reader interest can be measured in terms of amount of reading done.

2. If two groups of readers read equal amounts of the very same story, interest of the two groups in the content may be considered equal. So far as these experimental results are indicative, however, the following tentative conclusions can be drawn:

1. Vocabulary difficulty is related to readership. Editors who are after greater readership should urge their writers to use words people understand readily—words which don't slow them down.

2. At certain levels of "human interest," readers are not attracted, and may, in fact, be repelled. Editors should caution their writers to avoid being overly "personal" in their copy when trying to get "human interest" into it.

3. In general, when interest in the content is high, "hard" words and "human interest" tend to have less effect on readership than content.6

It is true that editors traditionally have advised their writers and copyreaders to use "short words and sentences"—thus recognizing the two most critical factors of readability, Lostutter7 points out. They have advised it but they have not always required it.

Foreign news, for example, is being written for readers with five years more education than the average adult American has, according to a recent study by Lester Getzloe.8 Lostutter states that it is not only foreign reports. Domestic news, for another example, presents such hurdles of hard words and abstractions to the average reader, who has gone to school less than nine years, giving

6 Ludwig, op. cit., 167-171
7 Lostutter, op. cit., 307-314
8 The Ohio Newspaper, Ohio State University, 28:2, 1946, cited by Lostutter, ibid., p. 308
the following example:

President Truman today called for undeclared war against totalitarian aggression to prevent further collapse of free institutions and further loss of independence in threatened countries...9

A recent investigation by Lostutter led him to believe that the newspaper world must recognize the need for readability on the same level as readership; that it must not stop at recognition but must go ahead to set up standards that will attain such a level of readability; and that it must then take a step further and apply the best available measurements to determine whether this readability has been attained—and is being maintained. Keeping a newspaper readable is about as much of a job as making it readable in the first place.

The State Journal, published daily in Lansing, Michigan, a city of approximately 80,000 population, was selected for his study. Lostutter feels that this paper with its circulation of 50,496 may be considered representative of papers of its own size and many much larger and much smaller, because it is big enough to have an editorial staff with some variety of background and specialization yet small enough to take the folksy attitude toward the news of the community that is characteristic of the non-

The metropolitan press of the United States. Such an operational definition of representativeness is in contradiction with the statistical meaning, the latter being widely accepted.

The prediction of readability requires "calculation by means of an empirical formula relating specific variables of readability to the criterion of readability."

The variables that have been used in Lostutter's study, in various combinations, are aspects of vocabulary, sentence structure and style, and reader interest. Criteria that have been developed include passages of graded difficulty from books and magazines, such as the 376 passages in McCall and Crabb's Standard Test Lessons in Reading and Flesch's 375 test passages from magazines "of five clearly distinguishable levels of difficulty," ranging from True Confessions to the Yale Review.

Lostutter believes that the formulas of Lorge and Flesch were best adapted to newspaper measurement for these reasons: (1) they are simpler than some of the others; (2) they are more recent and embody the experience of the

---


11 Rudolph Flesch, "Marks of a Readable Style," Teachers College, Columbia University, 1943
the earlier investigations; and (3) they take adult reading into consideration. However, says Lostutter, the newspaperman will find the Lorge formula tedious in three respects—marking the hard words, determining how many are different, and making the computations. He found that the Lorge measurements took him two to three times as long as Flesch's.

The conclusions to be drawn from his investigation are:

1. If the State Journal is representative of American journalism, more measurement of the readability of newspapers is needed, both as a preliminary to campaigns for improvement and as a check on the results of such campaigns.

2. The Flesch formula is the one in current use best adapted to the measurement of newspaper material.

3. Closely allied to the second, is that more study is needed of the effects of personal names of newspaper readability. Do they constitute a critical factor? If so, have the authors of the existing measurement formulas taken this factor into account sufficiently and interpreted it correctly? Whatever the answers to those questions, however, it is a foregone conclusion that most editors will stick to their policy of using names, lots
of them, to gain readership.

4. Attainment of readability for the newspaper as a whole is a conscious process somewhat independent of the education and experience of the paper's staff writers. It has been seen that the reporter with a college education may write easier copy than one with a grammar school or high school background--or may not; that the veteran of twenty years' experience may write easier than the newcomer--or harder.

5. The newspaper lead, particularly the summary lead, needs attention and even some modification if more readability is desired. The practice of trying to answer all the "W's" in one breath is unnecessary. It rests partly on habits and traditions that have not undergone sufficiently the ordeal of analysis.\(^{12}\)

Swanson\(^{13}\) conducted a split-run experiment in July 1948 at the State University of Iowa. A story was written reporting two campus surveys on questions which student voters said they were asking about the presidential candidates. The story included questions which three professors of political science advised voters to ask.

\(^{12}\) Lostutter, op. cit., 307-314

\(^{13}\) Charles E. Swanson, "Readability and Readership; A Controlled Experiment," Journalism Quarterly, 25:339-343, 1948
This material was not reported by any other news medium.

The two classes of readers were so similar in characteristics that, if any difference was found in the number of paragraphs read, the difference must be attributed to a difference in the readability of the two articles. Swanson substantiates the foregoing by employing T-tests.

The article was rewritten in two versions, one with 173 syllables per one hundred words. The easier version also had certain other factors associated with ease of reading, and used in one combination or other by five readability formulas.

Other results were:

Total paragraphs read--The 173-syllable sample of 90 men read 1,191 paragraphs. The 131-syllable sample of 95 men read 2,301 paragraphs. This was a gain for the easier version of 93-11 per cent.

Mean number of paragraphs read--For the 173-syllable sample the mean number of paragraphs read was 13.077 paragraphs. For the 131-syllable sample the mean was 23.969 paragraphs. This was a gain for the easier version of 83.16 per cent.

Number of respondents reading every paragraph--Thirteen subjects in the 173-syllable sample and twenty-five in the 131-syllable sample read 58 paragraphs. This was a
gain for the easier version of 82.35 per cent.

Statistical analysis showed that the difference between the medians in the number of paragraphs read by the two samples was significant far beyond the one per cent level of confidence (probability equalled .0014).

So far as these results are indicative, then, the provision of increased readability appears to make for the readership of a larger number of paragraphs. Where writers for mass media can apply the principles of readability tested in this experiment, they can expect some increase in readership of their words in print.

Before presenting a closer look at readability formulas, few of the current techniques, under the heading of a concept called condensation, will be presented.

Effective news condensation—the art of telling the essential factors more clearly with fewer words—not only saves space but also processing and reading time for each story, says Martin.14 More condensed writing may be a better way to maintain an adequate flow of information under such conditions than merely throwing away whole stories or reducing type sizes.

Of course condensation has drawbacks. Achieving

maximum conciseness not infrequently requires more time than deadline-pressed newsmen can give. But the condensing gains speed as techniques become habit.

Over-condensation also can produce dense, dull, inadequate stories. But in proper balance, condensation can bring the news into sharper focus. "Conciseness is the essence of clarity; every unnecessary word is an obstacle to the transmission of thought," a writer has observed.¹⁵

Surprisingly, the exact methods of condensing are practically ignored in readability formulas, as well as in most of the 30 newswriting and editing textbooks searched for condensation techniques.¹⁶

Condensation seeks to connect the reader with the news by the shortest possible line of words. That line is shortened by three main processes: Omitting superfluous details, integrating related facts, and wording those facts economically. The first process is often called "cutting" or "trimming;" the third--"boiling down."

The following techniques are offered as guides, not as inflexible rules:


¹⁶ ibid., pp. 337-340
(A) Omit superfluous details.

Any Who-What-When-Where-Why-How details not essential to conveying the significant fact and feel of a story may be left out. These may include superfluous identification or location, over- attribution to authority, or minor and irrelevant actions.

(B) Integrate related facts to avoid duplication.

(1) In the over-all story organization, loose construction—particularly below the lead in inverted pyramid stories—wastes space through repetitious flashbacks and tie-ins. Related details usually can be pulled together in topical units or, sometimes, in chronological sequence.

(2) Three integrative space-savers can be used in sentence construction:

(a) Avoid following an indirect quotation with a virtually redundant direct quotation.

(b) Use of the "series" form lets the subject—the engine of a sentence—pull smoothly a longer train of closely related facts, thus saving space otherwise required for duplicating subjects and predicates and presenting minor details quickly.

(c) Interpolations also economize and answer reader-questions just when they arise.
(C) Word the essentials economically.

After culling and integrating have eliminated superfluous material, the round-about and the redundant, then economize on wording. Savings in grammatical form are minor in a single instance but considerable in the aggregate. It is usually better to use:

(1) Short words for long whenever the short one is just as precise ("buy" for "purchase" or "car" for automobile).

(2) Active voice for passive.

(3) Pronouns alternated with nouns when the antecedent is unmistakable.

(4) "How" verbs for (a) adverbs or (b) verbs which tell only "what" and require modifiers ("Chapman raced to first base" instead of "Chapman ran swiftly to first base.").

(5) Single words for prepositional phrases:

   (a) Infinitives ("Replacements were trained to use bazookas" instead of "Replacements were trained in the use of bazookas.").

   (b) Possessives ("England's problems" instead of The problems of England").

   (c) Adjectives ("A blue-eyed girl" instead of "A girl with blue eyes").
(d) Adverbs ("They assembled quietly" instead of "They assembled with little commotion").

(6) The specific rather than the general ("Churches distributed 27 food baskets" instead of "Churches distributed more than two dozen food baskets").

British newspapers provide the best modern examples of extreme condensation. Since 1940 they have struggled daily to compress the abundance of world news into four, six or eight pages. British metropolitan dailies, with a few notable exceptions like the London Times and the Manchester Guardian, elected to reduce the number of pages and try to present something about many items rather than more about fewer.

Several areas for further research in condensation suggest themselves. A major one would be the relationship between condensation of an article and the Flesch readability formula. The hypothesis would be: Articles written to a satisfactory degree of condensation would have a greater ease of reading score. Further, comprehensibility would be greater in an article "condensed" than a "non-condensed" article with a high ease of reading score by the Flesch formula.

Condensation's bibliography is small. But the value of knowing how to condense needs to be increasing with
growing pressure on newspaper space, with the development of facsimile and quick-type news magazine, and—perhaps not too far off—with the advent of more and smaller papers printed by new typographical methods.17

The Flesch formula got a mixed reception from men and women in journalism, says Kearl.18 Reporters, columnists, and others whose work is primarily writing frequently have been doubtful and sometimes open resentful at the suggestion that writing style can be given an arithmetical score. Their complaints vary. A columnist says that writing to a formula stifles his creative spirit. A city hall reporter may explain that since his editor read The Art of Plain Talk, the kick has gone out of

17 Some cited by Martin, ibid., p.340 are:
Joseph G. Herzberg, Late City Edition, New York: Henry Holt and Co., well stated points are scattered about in chapters on "News Writing" and "Rewrite."
Robert M. Neal, Editing the Small City Daily, New York: Prentice-Hall, Inc., 1940), is probably the most complete, devoting a chapter to condensation.
Norman J. Radder and John E. Stempel, Newspaper Editing, Make-up and Headlines, New York: McGraw-Hill Co., Inc., 1942, also has a relatively thorough discussion.
George A. Brandenburg, "Shorter Words and Stories Become Chicago News Policy," Editor & Publisher, 78:8, April 7, 1945.

writing. Nobody asks, "Is it a good story?" The only question is what kind of a Flesch score it gets.19

Editors, publishers, and teachers have been somewhat more receptive, Kearl goes on.20

In 1948, a great year for readability formulas, these developments were observed:

Four major readability formulas were discussed and used.

Half a dozen universities conducted research to sharpen the tools of readability measurement.

A large number of newspapers and magazines using readability formulas, called upon readability experts to work on writing styles.

Kearl feels that some newspaper people's intuitive skepticism is subject to questioning. Many of their attacks against formulas are not made on rational grounds—but that these attacks hint at some real, and too seldom acknowledged, limitations of readability measurement. The salesmen of readability service—editors and teachers—try to carry readability scoring much farther than its research has justified.

19 Kearl cites, ibid., pp. 344-348, some—but by no means all—of these arguments are disposed of in Robert Gurney's short article on misconceptions about readability, Editor & Publisher, Sept. 13, 1947, p. 38.
20 ibid., pp. 344-348
Three formulas which received most attention during this period were as follows:

1. The original Flesch formula.

2. The revised Flesch formula.\(^{21}\) It's really two formulas: one for human interest and one for reading ease. Reading ease is determined by the length of words and the length of sentences, as in the earlier formula.

3. The Dale-Chall formula. It is concerned only with sentence length and proportion of hard words. Whether a word is hard or not is determined by whether it appears on a list of 3,000 words well known to a test group of fourth-grade pupils.

Each of the three formulas deals only with writing style. But obviously much more than writing style is involved in readability. Waples, Bradshaw, and Berelson suggest a list of elements of readability which includes, among other things:

- The climate of opinion in which a piece of writing appears.
- The Character and purpose of the writer.
- The format of the publication.
- The reader's expectation—whether he expects the item

\(^{21}\) Flesch, op. cit. pp. 221-233
to be dull or interesting.  

Such elements, continues Kearl, clearly have their effect on readability, but naturally are not touched upon by the equations generally described as readability formulas. Gray and Leary worked out an even more systematic summary of readability. They concluded that four major influences determine whether writing is readable: content, format, organization, and style. A second major limitation stems from the first. Even within their proper sphere of writing style, the formulas do not attempt to be complete or inclusive.

A statement with which Rudolf Flesch announced his first readability formula is of interest here. He quoted Lyman Bryson, director of the Columbia University readability laboratory, as saying that the term readability includes three aspects of reading matter: comprehensibility, lucidity, and appeal. Flesch explained that comprehensibility was the one of those qualities with which his work dealt. Thus at the outset he consciously ruled out two important elements of stylistic readability, in order to get a more accurate measure of the one that remained.

22 Douglas Waples, Bernard Berelson, and F. R. Bradshaw, What Reading Does to People, Chicago, 1940 pp.135-157

Admitting that there are many measurable characteristics that affect the comprehensibility of what you read, Flesch said, in effect: "Three characteristics are the most reliable and the easiest to measure. They are sentence length, word length in terms of prefixes and suffixes, and references to people."

The modern formulas, states Keen, are not unique in this approach. Herbert Spencer, Minto, Sherman, Bleyer, Stormzand, Rickert, Thorndyke, Dale, Gray, Leary, and Long all have experimented with this idea of style. But in each case the approach to the matter has been the same. Each researcher has sought to learn:

What elements of style are measurable;

Which, when measured, will give the best estimate of the quality of writing style.

Keen mentions a distinction made by Sherman which has been lost in the modern readability formulas: "Short sentences are not necessarily easy to read because they are short. Nor are long sentences always heavy and obscure because they are long." A long sentence, Sherman goes on, is not so bad if it has a few verbs in it, but a long sentence made up of dangling phrases is hard to follow.
In recent years, one very broad attempt has been made to determine how many different elements of writing style can be measured and analyzed. Willard G. Bleyer, for his University of Wisconsin style seminar, totalled up nearly 200 measurable ways in which writing styles could vary. But, of course, there was no intention that 200 elements of style be included in a formula. Each proposed formula has been a simplification of this list. In other words, the Flesch formulas and the Dale-Chall formula have deliberately confined themselves to two or three characteristics of style, not because only two or three out of 200 are important, but because a line must be drawn somewhere between completeness and practicality.

The result is the stylistic elements that have been omitted can easily outweigh those left in. Consequently, writing done at the command of the formula will not necessarily be good writing. It may not even be readable or understandable.

Two other minor points are important in considering the use of the Flesch and Dale-Chall formulas in newspaper work. The first is that, by their mechanical inflexibility, they may encourage extremes. In the eyes of the formula there is no ideal average sentence length: the shorter the better. Similarly, the shorter or simpler the words

---

25 cf. ante, p. 22
used, the better the score. Failure to see that these better scores do not always mean better writing inevitably yields a primer style.

The second minor point, which also needs to be recognized when newswriting is considered, is that all three formulas give an unconscious advantage to material containing dialogue or conversation.

Whether newspaper writing should make more use of dialogue, or whether an entirely new style should be invented which, without great use of dialogue, is nevertheless personal and varied in pace, is an open question. But this much seems fairly certain: unless the newspapers take the one step or the other, they cannot expect to reach the lower registers of readability as measured by the three formulas in current use.

Readability research has been highly commercialized in the last few years. In many ways this has been a very good thing; in others, important aspects have been ignored because they were not profitable to investigate. Newspapers and press services have become reader-conscious to a degree unprecedented in modern times. But because readability formulas have been commercialized, often much more attention has been focused on their potentialities than on their limitations.
A study by Griffin deals with an experiment with two versions of a news story.\textsuperscript{26} The original version (Story A) appeared in the \textit{San Francisco Chronicle}; the rewrite (Story B) was prepared by a journeyman reporter of the \textit{Chronicle} regarded as a competent craftsman. The changes from Story A to Story B were made on the basis of the professional judgment of the paper's news editor, with suggestion from Griffin, and in accordance with normal newspaper practice. Story B developed a different lead than did Story A; the structure was rearranged and the organization tightened.

Two populations were tested as to their differential in comprehending Story A and Story B. The first population, U-1, consisted of the 587 members of an organization of Berkeley women having a non-partisan, societal interest in government; the second population, U-2, consisted of the 253 members of two organizations of business and professional men having a non-partisan business interest in government. Comprehension was measured by three questions testing the immediate recall of essential elements in the story. Half of the interviews obtained were from women and half from men. Half of the respondents had attended college and half had less formal educational experience.

\textsuperscript{26} Philip F. Griffin, "Reader Comprehension of News Stories: A Preliminary Study," \textit{Journalism Quarterly}, 1949
In each of the exploratory studies the data collected fell into similar patterns and this was especially true of the data resulting from unanticipated responses. Objection to word usage was principally made in reference to the use of initials, names and special or technical language. When objections were made to sentence structure, the objection was more likely to refer to sentence's syntactical organization than to its length in words, or to the monotony of sentence rhythm. But certain design practices having traditional acceptance in newspaper craftsmanship drew the greatest total protest and seemed responsible both for failure to comprehend and for loss of reading interest.

These were specifically, the construction of a story according to the pattern of descending importance of the elements of the subject narrated, and the inclusion of considerable detail. The first objection was especially great when the pattern of the story required the reader to refer to a preceding paragraph in order to provide continuity to the composition. The second objection was most pronounced when names and facts were crowded into a story without, in the reader's judgment, their sufficient integration with the subject.

Finally, the effectiveness of newspaper stories as communication, as this effectiveness was measured by the
respondent’s ability to repeat the content of stories in sufficient and accurate detail, was disappointing.

The distinction between a bias and a projection as used in this study is that bias represents a precondition which was elicited by the story but does not identifiably interfere with recall. A projection is a product of a precondition and of the story and is revealed by the interjection of opinion into the respondent’s conception of the facts stated in the story.

Thus it may be presumed that a newspaper story which is better comprehended and read with greater interest may stimulate bias, but not against the general preconviction of the reader. However, maintains Griffin, projection plays a less desirable and more dangerous part, as well as one harder to identify. Thus it appears that the presentation of a more easily comprehended newspaper account which is more likely to sustain the reader’s interest does seem to increase the probability the projection will occur. But evidence that the degree or direction of the projection can be apprehended will have to be the result of further inquiry, says Griffin.

As a newspaper story is more readable and more greatly sustains interest, it becomes more incumbent upon the communicator to safeguard against language symbols that may
stimulate distorting projection and opinion-confirming bias.

He was not, says Griffin, able to establish any firm proof that the reader's attitude toward the newspaper might be affected by his awareness that the story he was reading was for him good (useful). Indeed, the evidence was that the reader will judge any story to be accurate and impartial as he believes it to conform to his preconceived attitude toward the subject at hand. But this judgment apparently will not necessarily affect his general judgment of the press in any way.

This study has clearly demonstrated, believes Griffin, that it is possible for a newspaper, under standard practice, to produce newspaper stories that will be more effective as communication. Further, this study has demonstrated that to some considerable degree newspaper stories which are not understood, are not understood because they are poor communication, and not exclusively because of faults in the audience.

In Griffin's judgment, one of the principal accomplishments of the study is the development of evidence that professional newspaper executives do not possess sufficient criteria upon which to judge the effectiveness of stories in terms of reader comprehension.
Griffin states that comprehension was measured by three questions. An operational definition of comprehension would have been in order while he was "operationally defining" bias and projection. It is questioned as to whether "comprehension" can be measured by three questions, especially when the author's expectation from the reader is that a sufficiently accurate comprehension of subject matter means to repeat the story's content in any considerable detail.

The two variables which Griffin considered was sex and education. There was no indication of an attempt to break down the levels of education. Just two categories are given: those who attended college and those who didn't. In his findings Griffin states that in no case did the test indicate that half the readers had a complete and accurate understanding of the story. No data are presented to indicate which half he has in mind: the college-attended or those with less than a college education, or what percentage from each group?

"Operational definitions" of bias and projection are presented. That latter seems peculiar to Griffin and is far from the basic meaning of the term. The term

27 Supra, 28

28 Supra, 30
"appereception" would have been appropriate. In any event, the employment of the concepts bias and projection are highly impractical in this study when one considers its purpose and scope. To be able to determine when a precondition which was elicited by the story but did not identifiably interfere with recall, or when there is interjection of opinion into the respondent's conception of the facts stated in the story, is not an easy task. Such areas of investigation require specific analysis, and such areas do not lend themselves readily to objective measurement.

Flesch reviewed the foregoing article and took exception to Griffin's statement that "he did not feel free to regularize the material tested by control of sentence length, word usage or story structure, since his purpose was to measure the effectiveness of professional performance in news presentation."

Flesch applied his revised readability formula and came up with the following readability analysis of Story A and Story B:

1. A rewrite prepared by a professional newspaperman without regard to the findings of readability research

---

29 Rudolph Flesch, "Reader Comprehension of News Stories: Further Comment," Journalism Quarterly, pp. 496-497, 1951
nevertheless raised the Reading Ease Score considerably. In other words the standards of readability research and of professional journalistic experience coincided in this case.

2. The difference in readability between Story A and Story B lay primarily in the degree of human interest. The increase in the Human Interest Score produced a very sizeable increase in comprehension. It also produced sizeable increase in approval of the story by the readers, as shown in other sections of Griffin's report. This finding conflicts to some extent with an earlier study by Ludwig in which an occasional adverse effect of high human interest was reported.

3. The direct relationship between the increase in readability and the increase in reader comprehension confirms the findings of several earlier studies. However, insofar as the test applied in this experiment may be considered as a test of recall, the findings reported here are novel. So far, no other studies of the important relationship between readability and recall have been published. (1951)

In reply to Flesch's "further comments" of Griffin's

31 Ludwig, op. cit., pp. 167-171
study, Griffin\textsuperscript{32} expresses his concurrence that the revision of Story A resulted in slight improvement of the Reading Ease Score and a more considerable improvement of the Human Interest Score. In his judgment, states Griffin, it is a matter of speculation as to whether these improvements are importantly related to the increased comprehension of Story B, the rewrite.

Griffin says that he is even less able to agree with the conclusion proposed in the second point advanced by Flesch. An examination of the data indicates that the respondents who read Story A were able to comprehend that story, because of the superior organization of Story B. Further contributing factors toward a superior comprehension by readers of B were that story's provision of the appropriate word and the well-constructed sentence. (The appropriate word was not necessarily the one of more common usage, nor was the effective sentence always of simple structure or shorter in words.

Because Flesch is primarily concerned with applying his formula to any and all studies, it is understandable why he did not criticize other aspects of Griffin's first article. Yet, it would have been within his province to point out some of the minor shortcomings of Griffin's article.

\textsuperscript{32} Philip F. Griffin, "Reader Comprehension of News Stories: A note on the Comment by Dr. Flesch," Journalism Quarterly, 28:497, 1931
Flesch would have been able to write a volume if he had reviewed an article by Fisk\textsuperscript{33} who made a study in an effort to ascertain and record possible stylistic differences between journalistic and "literary" writing. Approximately fifty thousand words of front page matter in thirteen leading daily newspapers were studied, and the results were compared with an equal amount of material from thirteen Literary Guild books of 1931-32.

Fisk gives no explanation for the basis of selection of the thirteen papers or the eleven books. She also included foreign writers. She did not differentiate between fiction and non-fiction, as well as whether the fiction is love, mystery, or adventure. She also omitted whether the same representative classes of people were used in each category.

After a presentation of much meaningless and insignificant data Fisk concludes with, "Perhaps, after all, the most striking conclusion derivable from this study is that there is less difference between journalistic and "literary" style than has been in some quarters supposed." The fact that six of the Guild authors used in her study were journalists was completely ignored. Additional criticisms of the validity of her findings have been omitted.

---

In 1943 Flesch developed a statistical formula for the objective measurement of readability (comprehension difficulty).\textsuperscript{34} The formula was based on a count of three language elements: average sentence length in words, number of affixes, and number of references to people.

Its validity has been reaffirmed by five independent studies, maintains Flesch: textbooks, advertisements, and media of communication.\textsuperscript{35}

The structural shortcoming of the formula is the fact that it does not always show the high readability deficit in conversational writing. For example, in the study of psychology texts, the score of Koffka's \textit{Principles of Gestalt Psychology} was 5.4 ("difficult"); yet William James' \textit{Principles of Psychology}, a classic example of readability, rated 6.0 (bordering on "very difficult"). Similarly, the formula consistently rates the popular \textit{Reader's Digest} more readable than the sophisticated \textit{New Yorker} magazine, although many educated readers consider the \textit{Reader's Digest} dull and the sprightly \textit{New Yorker} ten times as readable.

Aside from that, Flesch goes on, the practical applications of the formula led to several minor misinterpretations.

\textsuperscript{34} Rudolph Flesch, \textit{Marks of Readable Style; A Study in Adult Education}, New York: Teachers College, Columbia University, 1943

Sentence length, for instance, is the element with the heaviest weight; it is also the easiest to measure. As a result, this feature of the formula is often overemphasized, sometimes to the exclusion of the others. On the other hand, the second element--number of affixes--seems often difficult to apply; users of the formula found this count particularly tedious and admitted to uncertainty in spotting affixes. The third element--references to people--raised no such questions; but it was sometimes felt to be arbitrary and the underlying principle was often misunderstood.

The revision of the formula presented in this study, says Flesch, is an attempt to overcome these shortcomings and make the formula a more useful instrument.

In reanalyzing the test passages, the following elements were used:

1. Average sentence length in words.
2. Average word length in syllables, expressed as the number of syllables per one hundred words.
3. Average percentage of "personal words."--The same element was used in the earlier formula. However, the opportunity was used to test a clarified definition, which made no significant difference in correlation.
4. Average percentage of "personal sentences."--This new element was designed to correct the structural
shortcoming of the earlier formula.

Formula A is for predicting "reading ease;" formula B is for predicting "human interest."

Formula A alone, with a correlation coefficient of .70, has almost as high a prediction value as the combined earlier formula whose correlation coefficient was .74. Formula B has a much lower correlation coefficient of .43 and, accordingly, does not seem to contribute much to the measurement of readability. It should be remember, states Flesch, that because of the criterion used, Formula B predicts only the effect of the two "human interest" elements on comprehension. In other words, the correlation coefficient shows only to what extent human interest in a given text will make the reader understand it better. The real value of this formula, however, lies in the fact that human interest will also increase the reader's attention and his motivation for continued reading, Flesch maintains.

It is strongly felt that "Human interest" as defined by Flesch is insufficient for establishing motivation for continued reading. Flesch exhibits his psychological naivete by making such a claim.

The significance of Formula A will be more easily understood when it is realized that the measurement of
word complexity and that word complexity in turn is indirectly a measurement of abstraction: the correlation between the number of affixes and that of abstract words was found to be .78. Similarly, the measurement of sentence length is indirectly a measurement of sentence complexity, Flesch maintains. Sentence complexity, in turn, may again be considered as a measure of abstraction. Formula A, therefore, asserts Flesch, is essentially a test of the level of abstraction.

As stated, the old formula rated the Reader's Digest significantly more readable than the New Yorker; the new formula A also shows that the Reader's Digest is significantly easier to read. But the new formula B clearly shows a large difference in human interest in favor of the New Yorker.

As readability formulas become increasingly popular, they must, of course, be evaluated critically, Hayes, Jenkins, and Walker36 maintain. Like other psychological tools they must be tested for validity and reliability.

In their first study, the authors found that some of the greatest discrepancies obviously appeared in inter-relations of personal sentences. A second source of disagreement was in determining reliability involving rhetorical questions.

A second study was conducted by the authors to test their findings with a large number of inexperienced analysts. The correlations between these two groups approximated those found in the first study and would lead one to the same conclusions. Reading ease with its components is analyzed quite reliably and human interest with its components is analyzed with less, though still fair, reliability. Analysis of personal sentences again shows the greatest lack of agreement between analysts.

The reliability of the original and the simplified Flesch reading ease formula based on (a) samples drawn from house organs, using thirteen pairs of relatively inexperienced analysts; and (b) samples drawn from books, using a single, more experienced analyst is reported by England, Thomas, and Paterson.37 The findings confirm the earlier reliability study by Hayes, Jenkins, and Walker38 and show that both the original and the simplified Flesch reading ease formulas are highly reliable. With heterogeneous materials, test-retest reliability coefficients from /0.95 to /0.99 were obtained. Flesch felt that if more heterogeneous materials were used the correlation coefficient would be lowered. Such a view, England, Thomas, and Paterson maintain, is naive, statistically speaking. Intercorrelations between the original

38 Hayes, Jenkins, and Walker, op. cit., pp.22-26
and simplified formulas are likewise "high."

As a continuation of his previous article Flesch states that students of communication agree that awareness of the level of abstraction is essential for full comprehension. For example, Perrin, in his college text, says:

"For exact and reasonable communication it is highly important that a speaker or writer knows where in the range of meaning of abstract words his core of meaning falls and that he makes this clear to his listeners or readers."51

To adapt the technique of readability measurement to the measurement of the level of abstraction, Flesch utilized the fact that certain parts of speech are more frequent in abstract expression, while certain other parts of speech are more frequent in concrete expression.

To sum up the hypothesis underlying the present study by Flesch: Level of abstraction can be estimated by computing the ratio of certain parts of speech to certain other parts of speech in written expression. Since level of abstraction is a basic element in readability (comprehension difficulty), this ratio can be used as a measure of readability, either by itself or in combination with other elements.

49 Flesch, op. cit., pp. 221-233


52 Flesch, ibid., p. 385
Flesch found that most parts of speech contained certain categories that were statistically related to abstractness and certain others related in concreteness. In general, words related to abstractness are more "indefinite;" those related to concreteness are more "definite."

The percentage of "definite words" appears to be a useful test in two ways:

First, it is a rough measure of the level of abstraction. As such, it may be used as a tool in semantic studies, critical reading, literary appreciation, translation, rating of advertising copy, and propaganda analysis. Step-by-step analysis of the level of abstraction in a given piece of discourse may also be helpful in logical analysis and the discovery of faults in reasoning, Flesch believes. Students of logic will disagree with such a contention. It is only through analysis of the arrangement of "definite words" will one be able to discover the faults in reasoning. Any syllogism can establish this point.

Second, the new test is a measure of readability, says Flesch. As such it replaces the two tests of "human interest" (percentage of "personal words" and "personal sentences") that formed part B of the author's earlier formula. Combined with the average word length in syllables, it gives a practical and comprehensive measure of readability.

It should be noted that this measure approaches the
problem of readability in a new way, since it emphasizes the importance of the newspaperman's "Five W's", the role of names and addresses, facts and figures, and the use of practical examples, illustrations, and anecdotes. Negatively, it points up the bad effects of too many adjectives, abstract nouns, and weak verb forms. It may, therefore, be expected to become a useful training device in the teaching of composition.

One author suggests that as a simple device for counting the number of definite words, the space bar on an ordinary typewriter is helpful. Set the margin stop at 0, and hit the space bar each time a definite word is encountered. When the sample passage is completed, read the total from the cylinder scales. To get the syllable count, go through the passage again, this time counting all syllables (except the first) in all words of more than one syllable, and add the total to the number of words tested.

Then consult his accompanying table to obtain corresponding readability scores. This table is similar to that prepared by Farr and Jenkins for Flesch's earlier formulas.

Farr and Jenkins devised two tables, one for "reading ease" or level of difficulty, and the other is "Human Interest."


These tables permit rapid and accurate determination of
the Flesch index values and eliminate all calculations
previously involved.

Flesch's article, "Measuring the Level of Abstraction," presents a sound empirical approach to the evaluation of
abstract words in a piece of writing, Jenkins and Jones⁵⁵ feel. They feel, however, that Flesch's application of the
findings to the measurement of reading ease may be some-
what misleading to the reader who believes that this formula
is necessarily better than the one previously presented.

Jenkins and Jones, observing an increase of .03 in
correlation coefficients in the measurement of three sets
of variables, question the significance of this gain,
either statistically or practically.

Flesch stated that sentence length and word length are
measures of abstraction and concluded, "formula A, therefore,
is essentially a test of the level of abstraction." The
authors, Jenkins and Jones, wonder to what extent Formula A'
and the latest measure of abstraction are correlated. If
this correlation were high, the earlier formula might pro-
vide a convenient shortcut to the more recent measure. Further,
they believe that the procedure for obtaining the count of
definite words seems to be much too complex for general use.

⁵⁵ James J. Jenkins and Robert L. Jones, "Flesch's
'Measuring the Level of Abstraction.'" Journal of Applied
Psychology, 35:68 1951
Essentially they feel that if one is interested in abstraction per se this new measurement may be of definite value. But as a measure of readability they feel that this offers very little to the worker in the field.

Such criticisms brought a quick reply by Flesch\textsuperscript{56} who at first expressed general concurrence that the new formula is not necessarily better as a measure of readability than the earlier one, that it is primarily a measure of abstraction level and only secondarily a measure of readability. Flesch further concedes that the count of definite words is a rather complex task.

However, in comparing it with the earlier formula, it should be remembered that it is offered as a substitute for Formula B rather than Formula A, he points out. As shown in the paper by Hayes, Jenkins, and Walker\textsuperscript{57} the analyst-to-analyst reliability of Formula B is somewhat lower than that of Formula A.

Flesch states that to evaluate the new formula solely on the basis of its numerical correlation with the criterion is to take too narrow a view of readability measurement. Jenkins and Jones\textsuperscript{58} seem to consider readability as a readily defined, fixed quality, so that the problem of measurement

\textsuperscript{56} Rudolph Flesch, "Reply to Criticism by Jenkins and Jones," \textit{Journal of Applied Psychology}, 35:69, 1951

\textsuperscript{57} Hayes, Jenkins, and Walker, \textit{op. cit.}, p.22-26

\textsuperscript{58} Jenkins and Jones, \textit{op. cit.}, p. 68
becomes simply the problem of finding the most easily applicable yardstick. However, Flesch goes on, readability surely can not be defined in terms of the criterion used—that is, the grade level of children who could answer correctly three-fourths of the test questions appended to McCall-Crabbs' test lessons. Rather, readability is a complex quality of written prose that relates to a variety of factors on the part of readers on different levels of age and education, for example, comprehension, readership, reading speed, recall, and attitude changes. In other words, in its wider context, a readability measurement formula should be considered as a diagnostic and clinical tool in the pathology of communication. Viewed in that way, the complexity of the new formula will have to be balanced against its clinical value. It is possible, for instance, that the count of definite words will give a better prediction of readership or recall than any other readability measures so far proposed. If so, the clinical value may justify the complexity of the tool just as, in another field of psychology, the clinical value has justified the complexity of the Rohrschach test.

It occurred to Farr, Jenkins, and Paterson59 that a further simplification of the Flesch reading ease formula

is possible by merely counting the number of syllables per one hundred words. The authors' justification for this purpose is based on a review of the work done in this area. They quote Johnson60 who proposed that the percent of polysyllabic words found in a thirty-one-hundred word samples per book could be used as a rough measure for determining the reading difficulty of elementary school textbooks.

In 1934, Dale and Tyler61 computed correlations between a large number of language factors. Percentage of one syllable words was shown to be as good an index or a better index than some eight other factors.

In 1935, Gray and Leary62 published their important and highly significant monograph. Percentage of one syllable words proved to be approximately as good an index of comprehension difficulty as did average sentence-length in syllables. An additional important finding in the Gray-Leary study is relevant here. By studying various combinations of four factors each by multiple correlation methods, they found that nine different combinations will each give an


estimate of difficulty which approximates that obtained by
the use of more elements. In other words, one quickly
reaches the point of diminishing returns, so that difficulty
can be predicted as well by a small number of linguistic
elements as by a larger number.

Study of 360 hundred word samples from twenty-two
employee handbooks shows a high correlation between these
two variables--Flesch's reading ease formula and number
of one syllable words per hundred words. Based on this
high correlation they obtained, a new reading ease regression
equation was derived and a table to facilitate computation
of the new reading ease index was prepared. This table,
the authors feel, gives a wider range of reading ease scores
than was true of the Farr-Jenkins tables.63

Both methods of counting word length correlated quite
highly with the reading ease scores. The correlation be-
tween the old reading ease scores and the new reading ease
scores for the 360 paragraphs was .93.

Finally, the average reading ease scores for the
twenty-two booklets computed on the basis of the old and
the new formulas was .95.

Flesch64 feels that the foregoing proposals by Farr,

63 Farr and Jenkins, op. cit., p. 278

64 Rudolf Flesch, "Reply to 'Simplification of Flesch
Reading Ease Formula'" Journal of Applied Psychology,
36:54-55, 1952
Jenkins and Paterson are open to criticism on two grounds:

1. The *raison d'etre* for the new formula is that it is a "simplification." Farr, Jenkins, and Paterson say that "this simpler method would obviously be much faster and would require no knowledge of syllabication on the part of the analyst." However, states Flesch, this is not obvious at all.

2. Farr, Jenkins, and Paterson based their new formula on 360 one hundred word samples from twenty-two General Motors employee handbooks. Flesch maintains that the correlation that holds for twenty-two homogeneous General Motors handbooks within the narrow range of 36 to 57 ("Difficult" to "Fairly Difficult") is apt to drop rather than rise when heterogeneous materials over a wide range of Reading Ease Scores are sampled.

The proposed new formula seems a step in an undesirable direction, resulting in a cruder rather than a more precise measure. In effect, Flesch believes, Farr, Jenkins, and Paterson reduce readability to the use of short sentences and one-syllable words. This is a simplification with a vengeance, giving fresh ammunition to those critics of readability measurement who dismiss it as a movement toward "baby talk" or "primer style." After all, says Flesch, the count of one syllable words was abandoned by students of readability some fifteen years ago. To return to it
now would coarsen the technique of readability measurement and impair its value as a diagnostic tool in the improvement of communication.

Klare raises some questions about the two advantages claimed for counting one syllable words over counting all the syllables in words. First, he objects to the statement of Farr, Jenkins, and Paterson, that the simpler method would require no knowledge of syllabification on the part of the analyst, and quotes examples to substantiate his objection.

Secondly, he questions whether the "simpler method" would be much faster.

A third objection is "would not each counting error be magnified, and reliability be decreased, by the new method?" Since writing contains more syllables than one syllable words, this would seem to be the case unless one could be assured that the analyst would make several syllable errors to each word error.

In reply to Klare, supra, and Flesch who have criticized their "Simplification of Flesch Reading Ease Formula" Farr, Jenkins, and Paterson plead guilty to careless overstatement that "it would not require knowledge of syllabification on the part of the analyst." What should be

---

substituted is a statement to the effect that this simpler method requires less precise knowledge of how to break up polysyllabic words into their components because such words are to be ignored.66 Even though the authors admit that the simpler method requires almost as much knowledge of syllabification as does the more complex method the chief advantage of the simpler method lies in the fact that it would obviously be much faster.

But both Klare and Flesch deny that the simpler method would be faster. Klare, after considering the relationship between one syllable word counts and the number of syllables per one hundred words merely asserts his belief that no significant amount of time would be saved. Flesch, using similar reasoning, also denies that the simpler method would be faster. He even goes so far as to argue that about forty per cent more work would be required when the simpler method is used. The authors quote, "As the great Chicago physiologist, Anton J. Carlson, was so fond of saying, 'Vass iss de evidence?'

Instead of using armchair reasoning, the authors made arrangements in October, 1951 to obtain the "evidence." The mean time in seconds for making the one syllable word counts and looking up the reading ease scores in the Farr,

---
Jenkins, and Paterson table was 82 with a standard deviation of the distribution of 36.8. The mean time for making the syllable counts and looking up the reading ease scores in the Farr Jenkins table was 147 with an S.D. of 62.8. Thus, the evidence fully substantiates the claim that the simpler method is obviously faster.

Klare believes that the simpler method would magnify each counting error and thus decrease reliability. A thorough-going study of the reliability of both methods would be needed to settle this issue. Data obtained by Farr, Jenkins, and Paterson, however, would not lead one to put much stock in Klare's belief.

Flesch attacks the authors' notion that $r$ between the old and the new reading ease scores will be higher for more heterogeneous materials than for the employee handbooks used in developing the simpler formula. He gives data for eleven samples from his How to Test Readability\(^{67}\) and claims that the new formula is not as sensitive as the old because the ceiling for one syllable words seems to be eighty and the floor about sixty. He therefore concludes that the new formula underrates both ease and difficulty.

Farr, Jenkins, and Paterson recomputed the data for the same eleven samples as a check on the reliability of the

\(^{67}\) Flesch, "How to Test Readability," (New York: Harper and Brothers) 1951, cited by Farr, Jenkins, and Paterson, ibid. p. 55-57
counting procedure and as a check on the claim that the new formula introduces a systematic bias at the extremes. It will be noted that no such bias really exists. Furthermore, although slight differences in the one syllable word counts are shown, no serious errors in the new reading ease scores are involved.

Flesch's claim that the new formula is not as sensitive at the extremes of difficulty is not borne out by the seven additional samples taken from the rest of Flesch's book. Data obtained by the authors show, in two examples, the new formula to be more sensitive than the old in the sense of yielding a lower reading ease score. Another example shows the reverse effect--the new formula yields a higher reading ease score.

Flesch's final criticism is that the new formula is a step in the wrong direction making reading ease formulas even more vulnerable to the change of encouraging "baby talk" and "primer style." The authors, too, deplore this type of charge because it is unfair but they believe that no large proportion of the "literary stylists" will attack the new formula than have been attacking the old formula on these grounds. Flesch, himself, has given effective answers to this "baby talk" type of attack in his How to

---

68 Flesch, How to Test Readability, cited by Farr, Jenkins, and Paterson, *op. cit.* pp. 53-57
Test Readability (4, pp. 40, 41, 45, 48, 49, and 50).

In conclusion, the authors would stress "time saving" as the great virtue of the new formula. It is to be hoped that this "time saving" virtue will lead to far greater utilization of Flesch's important contribution in a greater variety of situations than is now the case. As matters stand, there is reason to believe that many practical people think that it takes an expert to make readability studies.

Getting into the highly sophisticated area of statistics Dunnette and Maloney69 have factor-analyzed the original and the simplified Flesch reading ease formulas. A factorial experiment was undertaken to study the effects of various factors on the accuracy and time taken by naive subjects to perform readability counts. The factors investigated were: (1) difficulty of reading material; (2) the type of count performed; (3) reading ability of persons performing the counts; and (4) sex.

The major finding was that the counting of one syllable words could be done in about three-fourths the time required for counting syllables. Boys performed the former counts more accurately than the syllable count. This difference was not statistically significant among the girls.

A significant interaction effect was found between

---

difficulty level and type of count. The syllable count was performed more accurately for easy material; the one syllable word count was performed more accurately for difficult material. Neither accuracy nor time taken was significantly associated with reading ability or sex.

It is therefore concluded by Dunnette and Maloney that the new Farr, Jenkins, and Paterson simplified Flesch reading ease formula is truly simplified since it can be applied with a greater degree of accuracy and requires less counting time.

The foregoing studies and criticisms of the Flesch readability formulas have generally been aimed at determining the reliability of them. In only one article the authors70 point out that "like other psychological tools they must be tested for validity and reliability." However, they concerned themselves with reliability. In "Validity of Readability Formulas" Swanson and Fox71 show that few validation studies using comprehension and retention as criteria have been made.

In their pioneering study Gray and Leary72 found twenty-four factors of style related to reading comprehension of adults. Gray and Leary, Dale and Chall,73 and Flesch

70 supra, p. 40
reduced these to a few factors. Their findings agreed on word difficulty and sentence length. Flesch also used personal references in one of his formulas.

In two experiments with articles in a midwestern farm paper Ludwig74 varied one factor at a time, word difficulty and personal references. His test articles were each read by more than forty per cent of the two samples of farmers. Readership differences between the experimental pairs of articles were small and were not significant.

Analysis of Ludwig's findings, state Swanson and Fox, suggested several hypotheses: Readability factors would have maximum effect when two or more positively related factors were varied. Easier words and shorter sentences, for example, should result in increases of comprehension, other things being equal.

Where more than forty per cent of an audience selects and reads an article, less gains in effect can be expected from improved readability. Also, where lesser proportions of an audience read an article, the more that gains may come from increases in readability.

Motivational factors inherent in content, such as subject matter, probably are more important, generally, than readability where individuals select what they want to read

74 M. C. Ludwig, "Hard Words and Human Interest," Journalism Quarterly, 26:167-171, 1949
and learn from printed media. For example, comic strips are easy to read but vary widely in readership, or audience interest. One comic strip may reach seventy per cent and another strip in the same day's newspaper reach twenty per cent of the same audience.

Readability factors might be more important than motivational factors where individuals are required to read and study and are tested on their learning. This would be the case in classroom and training situations.

In this study by Swanson and Fox, easier and harder versions of twelve articles were published in three issues of a paper sent monthly to employees of a mid-western company. Four articles appeared each month. The 296 employees were randomized into two groups, "easy sample" and "difficult sample."

Easy sample received copies of the newspaper with easier versions of the twelve articles. Difficult sample received the same newspaper with the harder versions.

Effects of the versions were determined by these criteria: (1) Retention, measured by a forty-three-item test of multiple-choice questions based on the twelve articles; (2) Readership, measured on easier or harder versions of two articles; (3) comprehension, measured by a ten-item test given before and after exposure to easier or harder versions of two articles.
Four other instruments were used. They involved general opinions about company union, general satisfaction with one's job, Sanford's authoritarian-equalitarian scale,\textsuperscript{75} and Goossen's distinguished intelligence test.\textsuperscript{76}

The following reliability differences appeared.

Formula scores. By the Flesch formula, the easier versions had a mean score of seventy-three (fairly easy) whereas, the harder versions scored an average of fifty-nine (fairly difficult). The Dale-Chall formula gives similar results. The easier versions had a mean Dale-Chall score of seventh to eighth grade compared with a score of eleventh-twelfth grade for the harder versions.

Number of words. The easier versions had fewer words, an average of 284, while the harder versions had an average of 322 words. The easier versions totaled 3,410, and the harder versions, 3,983 words.

Flesch Human Interest Index. The easier versions had a mean score of forty-six (very interesting) and the harder versions a mean of seventeen (mildly interesting).

Sentence length. The easier versions had an average sentence length of thirteen words; the harder, nineteen point four.

\textsuperscript{75} F. H. Sanford, \textit{Authoritarianism and Leadership}, Philadelphia: Stephenson Brothers, 1950 cited by Swanson and Fox, \textit{op. cit.} pp. 114-118

Syllables per one hundred words. The easier versions had 142 syllables per one hundred and the harder versions, 161 syllables per one hundred words.

Unfamiliar words. As scored by the Dale-Chall list of 3,000 unfamiliar words, the easier versions had 11.6 per cent unfamiliar words whereas the harder versions had twenty per cent unfamiliar words.

Verbs and adjectives. The easier versions had 130 verbs per one hundred adjectives. The harder versions, eighty-nine.

This study could not answer the question of whether some of these or other reliability factors cancelled out comprehension gains. No previous research had been published at the time of the designed study to suggest this possibility. The Gray-Leary and Swanson investigations indicated that reliability factors such as these would combine for positive effects.

Results.

Readership. The two samples did not differ significantly in readership. Of the easy sample, sixty-five per cent (n=67) read both articles; of the difficult sample, sixty-one per cent (n=63) read both articles.

Subject: exposed to harder versions succeeded as well on a forty-three-item information test as those exposed to easier versions.
Subjects who read easier versions of the two articles in a test situation did significantly better on a ten-item test of comprehension than those who read harder versions. This result indicates that readability formulas can predict some differences in comprehension between versions of the same material.

Readers of two articles were more successful on the forty-three-item test of information in the twelve articles than those who had not read either of the two articles tested for comprehensibility.

While these results indicate that readability formulas can be used to predict differences in comprehension between two versions of the same material, the findings do not support the utility of such formulas in predicting differences in readership, and retention for similar material, conditions, and time periods. Even combined treatment of readability factors, such as was attempted in this study, did not influence retention.

One factor limiting these results is the relatively high interest (readership by sixty per cent of the samples) in two of the articles.

The lack of differences in retention between easier and harder versions suggests that investigation of motivational factors inherent in content is most crucial where individuals select what they want to read and learn. This does not gainsay
the possibly greater importance of readability where individuals are required to read and study as in classroom and training situations.
CHAPTER III

STYLES OF TYPE FACES AS FACTORS OF READABILITY

As part of an extensive investigation of leading and line width in relation to type size, Paterson and Tinker\(^1\) studied the influence of leading upon readability of eight-point book type printed in seven, fourteen, twenty-one, twenty-eight, and thirty-six pica line widths.\(^2\) The results revealed that no significant changes in readability occurred when line width and leading varied over a rather wide range. The purpose of the present study is to determine the effect of varying line width and leading upon the readability of eight-point newspaper type.\(^2\)

The choice of the variables was guided somewhat by the results obtained by the authors in their survey of printing practice for newspaper body type.\(^3\) This survey revealed that Excelsior type was more than holding its own over a period of years, that the use of eight-point type was rapidly increasing, that a twelve-pica line width was almost universal on the front page of newspapers, although line widths up to twenty-five and one half picas were employed on

\(^{1}\) D. G. Paterson and M. A. Tinker, How to Make Type Readable, (Harper & Brothers), 1940, pp. 72-81


\(^{3}\) D. G. Paterson and M. A. Tinker, "War Time Changes in Newspaper Printing Practice," Journalism Quarterly, 21:7-11, 1944
editorial pages, and that set solid, one half, one, and two point leading were most common.

The authors used a speed of reading technique to measure readability. The material consisted of Forms A and B of the Chapman-Cook Speed of Reading Test. In all comparisons Form A was printed in eight-point Excelsior type face, twelve-pica line width with two-point leading on newsprint paper stock. Form B was eight-point in six, twelve, eighteen, twenty-four, thirty, and thirty-six pica line widths. Each line width was set up with zero (set solid), one-half, one and two point leading. Each variation of line width and leading of Form B was compared with Form A, making twenty-four comparisons in all. Twenty-four groups of eight-four high school seniors were tested, or 2016 in all.

In general, the data indicate that for eight-point Excelsior type:

1. An eighteen-pica line width with one or two point leading produces most readable text.

2. Textual materials with a rather wide range of line widths and leading are equally legible.

3. Very short and relatively long line widths produce poor readability except when the long lines are generously leaded.

Readers judge text with the larger amounts of leading to be more legible and to be more pleasing for each length
of line considered. Therefore, when a twelve-pica line width is used for eight-point type, one or two points of leading should be used if reader opinions are considered important, even though the leading does not increase readability.

In 1936, Paterson and Tinker made a survey of styles of type faces used in front page body type composition and on editorial pages of eighty-nine newspapers, large and small, published throughout the United States. This survey revealed only fifteen different type faces in use. They selected the following seven most frequently used type faces for study:

1. Ionic No. 5, Ideal, Excelsior, Regal No. 1, Century Expanded, Textype, and Ionic No. 2. In addition, two relatively new type faces (paragon and Opticon) were added.

The Chapman-Cook Speed of Reading Test was employed to measure the relative readability of the nine type faces. In all comparisons Form A was printed in Ionic No. 5 type face as a standard. Form B, which was read after Form A, was printed in one of the nine faces. Nine groups of one-hundred subjects (high school seniors) each were tested.

It was found that the three most readable type faces are Opticon, Regal No. 1 and Century Expanded. Text set in

---

4 D. G. Paterson and M. A. Tinker, "Differences Among Newspaper Body Types in Readability," *Journalism Quarterly*, 20:152-155, 1943
these three type faces are read 6.1 to 7.8 per cent faster than the standard. Paragon, Excelsior, and Ideal group themselves as being read significantly faster than the standard (4.6 to 5.6 per cent). Ionic No. 2 and Textype are read slightly faster than the standard but the differences are not statistically significant. In this study, using 900 adult readers, it happens that Ionic No. 5, the standard, was read more slowly than any other type face.

Still using the Chapman-Cook Speed of Reading Test as a measuring rod, Paterson and Tinker conducted another study to determine the influence of leading on speed of reading newsprint and apparent legibility. Eighty-five university students in each group, a total of eight groups, were tested, as well as an additional group of 225 students.

It was found that text set solid was read slower than all other materials with leading. Significant differences beyond the one per cent level were found between text with four and five point leading and set solid text, and nine-point leading was judged less effective than most other amounts of leading.

In another study Paterson and Tinker attempted to determine the influence of line width and leading on the

5 D. G. Paterson and M. A. Tinker, "Speed of Reading Nine-Point Type in Relation to Line Width and Leading," Journal of Applied Psychology, 33:81-82, 1949

speed of reading nine-point type.

The results indicate that the optimal rate of reading occurs when line widths of fourteen to thirty picas and with one to four point leading. This may be considered the zone of safety. A conservative range would be sixteen to twenty-four pica line width with one or two points leading when nine-point type is used.

Moving into the area of readability of newspaper headlines printed in capitals and lower case, Paterson and Tinker, in the first part of their study,\(^7\) dealt with five word single-column headlines set in twenty-four point bold face, half in upper case and half in lower case, to be read by tachistoscopic exposure at the normal reading distance of fifteen inches. The results disclosed an 18.9 per cent difference in favor of the lower case headlines.

The third part of their study used multi-column or banner headlines set in sixty point bold face which were exposed at varying distances from six feet to seventeen feet. At a distance of six feet, the legibility of the lower case banner headline was 5.3 per cent greater than the legibility of those set in upper case. At distances from ten to fourteen feet, both kinds of headlines were equally legible. At a distance of seventeen feet, however, the upper case headlines proved to be more legible.

---

Breland and Breland did a similar experiment in newspaper headlines printed in capitals and in lower case. Each of 120 five-word headlines was printed once in capitals and once in lower case. Two sets of tachistoscopic slides were organized. In one set the first headline was in capitals and the second headline was in lower case, etc.; in the other set the first headline was in lower case and the second in capitals, etc.

Twenty-two subjects were presented the slides for fifty milliseconds each and asked to reproduce the headlines. Scoring was in terms of number of words correctly produced, omitting the first twenty trials. The scores indicate statistically significant differences in favor of greater legibility of headlines printed in lower case.

In an experiment for determining the illumination intensities for reading newspaper type, Tinker administered individual tests to 405 university sophomores to determine reading speed under each of seven levels of illumination, i.e., one, four, seven, ten, twenty, fifty, and one hundred foot-candles, and to compare efficiency under these conditions of lighting with records of a preliminary test given under

---


9 M. A. Tinker, "Illumination intensities for Reading Newspaper Type," Journal of Educational Psychology, 34:247-250, 1943
ten foot-candles.

The critical level for reading seven-point newspaper type, set solid, was found to be approximately seven foot-candles. At lower levels of illumination, speed is retarded; at higher levels, it is not significantly increased. To provide a margin of safety, the author suggests that fifteen to twenty foot-chandles of light be employed for newspaper reading.

For the second time in the present century, the physical format of daily newspapers is undergoing changes which in large measure, may be attributed to the need to reduce production costs, especially expenditures for high-priced newsprint. Oldtimers will recall the earlier nationwide change in format started slowly before World War I when, as an economy move to meet increasing newsprint prices, many daily newspapers changed from thirteen to twelve and one half or twelve pica columns. Following the war, a movement was started to narrow columns to a standard of twelve picas. In that change the metropolitan papers were in the vanguard, and were followed in turn by non-metropolitan daily newspapers, and later by the weekly press. In the thirty odd years since the start of the change, nearly all daily newspapers and about four-fifths of the weeklies reduced column measures from thirteen to twelve picas. Some dailes changed to narrower measures.
Employing sophisticated statistical methods Barnhart and Jones\textsuperscript{10} conducted a study to determine reader reaction to nine-column newspapers and page shrinkage.

They found that newspaper readers have great difficulty in discriminating any differences between $7/8"$ and $1 1/16"$ shrinkage of pages when other page format variables are constant. More than two-thirds of the sample in this study refused to state a preference between pages differing only in shrinkage. It appears that minor variations in page shrinkage will go unnoticed by the great majority of newspaper readers.

Results indicate, among those respondents who stated preferences, that: (a) eight-column front pages are preferred over nine-column front pages by a ratio of about two to one; (b) eight-column inside pages are preferred over nine-column inside pages by a ratio of about three to one. A substantial number of respondents, however (about one-third of the sample), expressed no preference for eight-column vs. nine-column pages.

Only about a fourth of the sample knew the number of columns in their daily newspaper. Knowledge of correct number of columns in their newspaper is not related to preferences for eight-column vs. nine-column pages. Breakdowns

\textsuperscript{10} Thomas F. Barnhart and Robert L. Jones, "Reader Reaction to Nine-Column Newspapers and Page Shrinkage," \emph{Journalism Quarterly}, 30:170-178, 1953
of the sample produced no systematic preference differences. Respondents did not appear to have strong feelings about their preferences. Many were much more concerned (a) that the newspaper be printed clearly (heavily inked) so as to assure good contrast, and (b) that the body type size not be reduced. They seemed to regard shrinkage and number of columns per page as less important than these two factors.

The principle of ocular photography (known as the corneal reflection technique) is valid as a means of obtaining an accurate record of eye movements. The camera employed photographs the eye movements in a bidimensional plane on a single film.¹¹ Findings of the Visual Research Laboratories have revealed that ocular patterns, resulting when reading, are indicative of the mental processes involved. Ocular patterns are as individual as the speech or walking habits of human individuals, the author maintains. However, certain characteristic eye movements are common to all.

CHAPTER IV
HEADLINES--THEIR FUNCTIONS AND PURPOSES

The function of the headline is to summarize in just a few words the most important elements of the story it covers. It conveys the highlights in short, easily understood, emphatic words. Strong verbs, colorful words, and the positive statement are characteristics. In effect, it is a miniature replica of the story and provides the reader with an opportunity not only to select at a glance what he cares to read but also to read the top of the news during the preceding twenty-four hours as he speeds for the 8:15 local.1

It is not the absolute size of the headline which is the determining factors as much as the position it has in the immediate or temporal ground; thus a large headline in a tabloid or in a Hearst paper does not necessarily impress the regular subscriber, because the front page of the latter paper has so many other large headlines and because both papers have similarly large headlines every day, not matter how significant or insignificant the day's news happens to be. The importance of a headline, moreover, is due to the fact2

---

that it evokes within the reader a definite attitude toward
the article before he has read a single line.

The newspaper reader is often a shopper of headlines,
so to speak, and may not allot sufficient time for a complete
reading of the headline writer's message. In one fixation,
says English, a reader may perceive quite distinctly the
symbols that fall on the fovea of the eye and less distinctly
those that fall around the fovea and on the periphery of the
retina. Thus in an attempt to perceive as much meaning as
possible in a hurried examination of a headline some of
the symbols may be seen less clearly than others.

It may be that certain type designs will furnish stronger
visual cues under these reading conditions than other dis-
similar designs, felt English at the outset of his experi-
ment, and thus increase the amount of meaning that is imp-
arted in a given time.

English states that while he wishes to measure head-
line display type from the point of view of "ease of seeing,"
the validity of the techniques available for scientific

3 Earl English, "A Study of the Readability of Four 1944
Newspaper Headline Types," Journalism Quarterly, 21:217-229

M. Luckiesh, "Some Comments on Dr. Tinker's Review of
'Reading as a Visual Task,'" ibid., 27:360-362. M.A. Tinker,
"A Reply to Dr. Luckiesh," ibid., 27:469-471 cited by
English, ibid., pp. 217-229
appraisal remains unestablished among the principal experiments in the field.

Some of the research methods discussed by English are: measuring visibility, measuring cost to the reader, measuring work output.

Measuring visibility. One approach to the problem introduces a consideration of visibility, a term referring to the intensity of the psychophysical stimulus which evokes perception and discrimination. Thus in a study of headline types one could "turn down the light" on type faces of different designs to determine which may be read with the least amount of illumination. For the measurement of the visual threshold, the point at which perception is barely possible, Luckiesh and Moss have developed a visibility meter. It consists of two identical colorless filters which may be rotated synchronously in front of the eyes of the observer, thus reducing or increasing the factor of brightness contrast. The degrees of density have been calibrated into two rational scales. One measures relative visibility, scale range one to twenty, and the other appraises foot.

5 Supra, p. 68


candles, recommended scale range one to 1,000. With the aid of this instrument Luckiesh and Moss have derived visibility ratings for an array of typographic conditions.8

Several related experiments have been done in this area, i.e., the visibility of print on various qualities of paper, the relative legibility of small letters, the relative legibility of black and white print, measuring of the "legibility" of letters by noting the distance from the observer that individual letters could be read, measuring the degree that out-of-focus letters could be identified when projected on a screen.

Measuring Cost to the Reader. In the search for a criterion that will accurately reflect ease of reading, one group of studies offers great promise—measures in cost to the reader. If some reading tasks are more difficult than others and the seeing mechanism is capable of responding to varying degrees of motivation, then it may be possible to seek out and measure evidence of degrees of fatigue arising from tasks of different difficulties.

Criteria of reading fatigue include metabolism and

pulse rate, variations in visual near point, eye movements, speed of accommodation, and rate of blinking. By far the most popular of these measures, probably because of the comparative simplicity of collecting data, is rate of blinking. Luckiesh and Moss maintain that the rate of the involuntary eyelid reflex increases as an axiomatically fatiguing visual task is prolonged. They report that it is necessary for a subject to read for only thirty to sixty minutes in order to show significant differences between the number of blinks in the first and last five-minute test periods. Among their consistently successful applications of this method is a study in which the relative rate of blinking for eighteen subjects increased from 100 in reading twelve-point type to 148 in reading six-point type.

The blink rate method has not been found as acceptable by other experimenters. After recent investigation at Harvard University, McFarland, Holway, and Hurvich reported:

The blink rate of some readers tends to increase as they continue to read. In others, it decreases with an increase of reading time. In still others it manifests no change at all. The blink rate, therefore, is not a valid index of visual fatigue.9

Measuring Work Output. In referring henceforth to all the intrinsic factors that affect ease of reading, i.e., visibility, visual effort, fatigue, etc., the more inclusive

---

term readability will be used, states English.

The readability of type faces has been investigated by measuring work output, or speed of reading, of a group of readers. This method has been favored by experimenters, particularly by Tinker and Paterson. They set the two equated forms of a reading test in different type sizes, designs or arrangements and administer them to a large group of readers. The difference in mean paragraphs read in the total time of 1 3/4ths minutes is then used as a criterion of readability.

While the consensus seems to be that this method is not ideal, Luckiesh and Moss have been particularly critical. They hold that the reading mechanism is of such an involuntary and complex a nature that it is virtually impossible to rely on speed of reading as a criterion of readability. They point out that in some cases a subject may respond to improved reading conditions by simply expending less visual effort.\(^\text{10}\)

So far almost all of the methods and studies mentioned have been concerned with body types. When display types have been measured, they have employed, with few exceptions, the criterion of visibility or visual fatigue as measured by the blink test. While high visibility is undoubtedly a basic factor in achieving high readability, it does not follow that

this consideration alone should guide selection of headline
types. For example, in a widely printed table of body type
rated as to visibility, six-point heavy Copper-plate capitals
are declared more visible than such lower case faces as
eight-point Caslon light, eight-point Cheltenham wide, and
eight-point Goudy light.\(^{11}\) The weight of experience and
research would hardly permit one to say that the same com-
parison would hold for the \underline{readability}\(^{12}\) of these faces.

The possibility of setting large display type into the
customary reading test forms to measure readability is quite
impracticable, English goes on. Furthermore, the exceptional
arrangement and purpose of newspaper headlines call for a
test that approximates actual reading conditions. On the
bases of the foregoing English decided to construct a special
headline reading test and to measure the number of words
read during uniformly brief exposure intervals.

English followed a rigorous procedure for the control
of any variables which would affect the validity of his
experiment. Other factors that remained uncontrolled are
differences in ability among the readers, and, of course,
the differences that may possibly lie in the effects of

\(^{11}\) W. Luckiesh and F. Moss, "The Visibility of Various
Type Faces," Journal of the Franklin Institute, 223:77-82,
1937, cited by English, \textit{ibid.}, pp. 217-229

\(^{12}\) As defined by English, supra, p. 77
design and size. English was cognizant also of other variables and attempted to take them into account. This was done by designing an experiment that yielded a dependable estimate of the effect of the variables not under control—a dependable estimate of error. With this estimate it was possible to determine the probability that the error of certain sources exceeded a certain magnitude and thus to gain an approximation of the maximum influence of the variables upon the results. He employs a method which lends itself ideally to these needs—analysis of variance.

Analysis of variance is a technique for isolating from comparable groups of data the variations traceable to specified sources. Thus his experimental design took cognizance of the basic assumption in analysis of variance that all extraneous factors which might introduce systematic differences have been randomized.

Further, the systematic error that might have arisen through use of a repetitive group order in presenting the type sizes and faces was invalidated by counterbalancing the order of presentation to the reader. As a result, each design and each size was presented an equal number of times in each of the three possible positions.13

13 Cf Leo Postman and James P. Egan, Experimental Psychology, (Harper & Brothers) 1949, p. 315
English found that:

1. Printing types under consideration in his study have been shown experimentally to differ in readability as defined.

2. The Cheltenham bold, Bodoni bold, and Tempo bold families were found superior to Karnak medium in the qualities measured in this test.

3. Cheltenham bold all-capsitals retard the speed with which headlines are read approximately eighteen per cent as compared with the lower case of this design.

4. There is reason to believe that the readability of headlines is fairly constant in the size range between fourteen point and thirty point, inclusive.

5. Reader judgment was found to be an unreliable index of the readability of type in headline form as measured by objective methods.

6. To increase the assurance that news presentation to readers occurs under optimum reading conditions further investigations of other headline type families seem justified.

Tannenbaum, in a study of the effect of headlines on the interpretation of the news,\(^\text{14}\) raises the question as to just what the role of the headline is in everyday newspaper

newspaper reading behavior. He feels that the headline is now an important, integral feature of the American newspaper page. Not only can the headline serve as an index by attracting attention to a particular item, but it may also serve as an index in terms of influencing the interpretation of a story, he says.

Most journalism textbooks warn against the introduction of partiality into headlines. All stress, like Taylor and Scher, that the headline "must leave no false impressions. It should not say less than the story, and it must not say more."

These warnings grow out of two factors in the process of communication by newspaper. First, the headline of today, by its very nature of extreme condensation and brevity, is often unable to present the desired "bird's-eye view" of the story beneath it. Space limitations make it often virtually impossible for the headline to tell the whole story. In most cases, then, the headline writer is forced to select a single aspect of the story to "play up" in the headline. Obviously, bias can enter here, virtually through an open door. Secondly, an even more important aspect of

---

15 Howard B. Taylor and Jacob Scher, Copy Reading and News Editing, (New York: Prentice-Hall, Inc.), 1951, pp. 137-194
how the headline can introduce bias stems from the way people read newspapers. The reader is essentially, as English puts it, "a shopper of headlines."\textsuperscript{16}

In 1928, Emig, using a simple questionnaire method, studied the source of opinions formed from newspaper reading. Of 375 subjects, 192 said that they based their opinions on "reading or skimming the headlines," 118 on reading news stories, and 144 on reading both headline and stories.

These data led Emig to conclude that headlines are "... perhaps the most potent factor entering into the formation and direction of public opinion in the United States." However, by Emig's own admission, neither the methodology followed nor the results obtained can be regarded as completely valid. The concept that headlines are a potent factor entering into the formation and direction of public opinion is highly questionable. During the late twenties the fallacy of the single explanation was a common one.

As part of a larger investigation into the ethics of American newspapers, Kingsbury and Hart and associates\textsuperscript{17} scaled headlines on the same topic from various newspapers, using the median of all newspaper reports as the base-line for comparison. On such issues as the militaristic-pacifistic

\textsuperscript{16} Supra, p. 73

problem, reparations, and prohibition—all vital problems at the time—bias was found in almost every case. However, granted that bias is present in headlines the problem as to whether this bias affects interpretation is still unsolved.

Much more to the point was the war-time study of Allport and Lepkin.18 These investigators attempted to find out whether headlines, as such, could arouse an attitude to "get in and do the fighting." The experimenters selected 126 "representative" headlines, reproduced them on cards in seventy-two point Gothic type with the same filler material running underneath. The cards were exposed for a twenty-second period to 109 "average citizens," who were asked to rate each headline on an eleven-point scale according to how they were made to feel toward participation in the war effort.

The results showed that all headlines—good news or bad—had some effect toward instigating a role in the war effort. The bad news headline, however, was significantly more effective toward this end than those which were classified as good news headlines.

This laboratory study, while illustrating the effect of the headline in predisposing the reader towards a

particular "feeling," still failed to duplicate ordinary newspaper reading behavior. The manner of presentation is hardly comparable to usual conditions, while it is not clear whether the subjects could read the actual filler copy.

Tannenbaum's study has demonstrated the fact that headlines are not impotent instruments in the formations of opinions from newspaper reading. Indeed, their effect appears to be a most profound one, even within the confines of this laboratory situation, the author feels. The results obtained cannot be explained solely by the properties inherent in the headline per se. As demonstrated by the results, the extent of reading is also an important factor contributing to the effect of the headlines. When we consider how much of ordinary newspaper reading is composed of headline scanning, with perhaps a glance at the lead paragraph of a particular story, we might well expect that the headline's effect would be maximized.

One other factor which may be expected to extend the influence of the headline in the usual newspaper situation arises from the nature of the experimental sample. While university undergraduates are not the most critical readers, they certainly do constitute a more sophisticated sample than one might expect from a representative segment of the total population. If this is so, an expectation of an even greater affect of the headline with a cross-population sample is not
unwarranted.

One must not forget, of course, that the body of the news story itself is probably still the main source of opinion and interpretation. But it appears that the headline sets the stage, as it were, for the manner in which the story is read, Tannenbaum believes. It establishes the frame of reference within which the factors of the story are perceived.

In the area of whether large headlines contribute towards sales most authorities in American journalism are in agreement. But today, when newspaper competition in many areas no longer assume major importance in distribution, how much does the headline contribute toward sales? The theory that large headlines promote sales apparently is still held in some journalistic and other circles. Albert A. Sutton expressed a similar view. In discussing social background of newspapers, Doob says: "To boost street-corner sales, the size of the headlines grew larger." Winship and Allport, however, after a three months study of what they termed seven representative newspapers,

19 Walter A. Steigleman, "Do Newspaper Headlines really Promote Street Sales?" Journalism Quarterly, 26:379-388, 1949
20 Bird and Merwin, op. cit. p. 169
concluded that there is no material effect upon sales of newspapers whether the headlines are "pleasant" or "unpleasant." Their war time study convinced them that newspaper sales are not affected by the "tone" of the headlines.

Steigleman states that with this study indicating the headlines has little direct sales appeal, with newsboys at times nullifying editorial judgment, and with readership surveys showing the streamer story does not always obtain maximum readership, journalists may well re-examine some present theories about the worth and function of today's headline. Front-page typography then could be designed, in the light of the function that headlines serve today, rather than the function they may have served in the past. Perhaps newspapers have been indulging in a wasteful and expensive ornament. The study of the effects of the headline, too, cannot ignore the findings of the continuing reader interest surveys that in a number of cases the top story in the judgment of editors does not always attract the highest reader interest.
CHAPTER V
READERSHIP: ITS COMPONENTS AND MEASUREMENTS

Few schools attempt to teach correct newspaper reading habits to early and middle teenagers, the age when they first become "readers" of the newspapers rather than just the comics section. One study suggests that the use of newspapers for reading instruction seems justified in junior high school classrooms. Quite a few teachers in Milwaukee have attempted to create an intelligent and critical interest in newspaper reading among their high school and elementary school students. Four of them collaborated for "The Newspaper in the Classroom" which is a tangible means of stimulating interest and effectively educating pupils to read newspapers.

A survey of an urban high school students' newspaper reading habits was done by Feingold. The group studied consists of 422 students during the prewar period. Feingold found that only three per cent did not read any newspaper, while sixty-nine per cent read more than one daily paper. There were no appreciable differences between the sexes in regard to method or plan of reading. The given reasons were: information, entertainment, and to pass time.

Eighty-five per cent believed their school work was benefited by the newspaper reading.

In another study of high school and college students over 400 protocols regarding likes and dislikes for newspaper reading were obtained from high school students in grades ten through twelve and from college students during their last three years in college. A list of thirty-five categories was prepared with such topics as photographs, comics, cartoons, and sports at one end and scandal, speeches, society, and advertisements at the other. The results show that illustrations and amusements appear to predominate among the reading interests of both high school and college students.

The relative appeal of newspaper items of twenty-nine different types was studied by direct observation for two weeks during the hours 7:30-9:00 a.m. and 3:30-5:30 p.m. of the reading of subway commuters and by a questionnaire given to 209 students of social science in grades one through twelve. Sports, cartoons, and photographs had the highest interest value in both groups. For the adult New Yorkers personal violence, disaster, and the serial story ranked high; radio, home-garden, church, and education, low. In the

4 P. Fendrick, "Newspaper reading Interests of High School and College Students," Journal of Educational Research 34:522-530, 1941
5 C. L. Ross, "Interests of Adults and High School Pupils in Newspaper Reading," School and Society, 27:212-214, 1928
student lists among the dominating interests were the theatre, city and national news, radio, and editorials. Divorce, indictment and jail; finance and trade, and education were the items with the least appeal.

Responses of pupils in grades eight through twelve to the thirty-six items of a specially designed attitude scale on the public press are compared with the reactions of Rotary Club members and school teachers to the same group of statements.6 Average scores of high school pupils suggest less favorable attitudes toward the press than those possessed by adults out of school, with evidence of a direct relationship between the number of years in school and the degree of hostility.

It seems evident that reading habits are developed in the early and middle teens says Wall in a study done in Britain.7 The attitudes formed toward the press in this period, however, are hardly serious or critical, Wall found.

Gerberich and Thalheimer submitted a questionnaire to college students and adults on newspaper interests in certain sections of a paper and in types of news stories.8 General


news, comics, editorials, and sports rank high in interest, while serial stories and art rank low. The news stories most preferred are travel and human interest. These are followed in descending order of preference by self-improve-
mant, sports, politics photographs, war, and education. The correlation in newspaper reading interests ranges between .50 to .90 for the various groups used in this study, although they differed in age and economic status. Correlations of interest between men and women range from .17 to .54.

The same authors devised a three-page questionnaire or reading report blank designed to elicit information from certain groups of newspaper readers concerning: (1) their interests in certain sections of the newspaper, (2) their interests in certain types of news stories, and (3) their attitudes toward certain questions of newspaper policy and practice.9 The mimeographed reading report blank was administered to samples of three population groups: freshman and upper class students of the University of Arkansas; business and professional men and women and housewives of Fayetteville, Arkansas, and vicinity; and journalism students of the University of Arkansas and various Arkansas high schools.

A lively, somewhat sensational treatment is indicated as the preference of a fairly significant majority; 64.6 per cent prefer large heads over small, 60.9 per cent prefer many stories with less detail to a few stories in detail, and 74.2 per cent prefer more rather than less human interest material. Especially significant to the small town paper is the 69.7 per cent preference for more than rather less local news.

Strong popular support of freedom of the press is indicated by the fact that 85.7 percent answered in the affirmative the question, "Do you think newspaper should be permitted to criticize the government?" This is by far the highest percentage of agreement reached on any item in this study. It is significant, however, that a large number feel that the press should be subjected to some kind of control, as indicated by a favorable response of 43.9 per cent to the question, "Do you think newspapers should be licensed by the federal government?"

Responses to questions involving the integrity of the press show that to a considerable degree it is the popular belief that the press is venal, biased and inaccurate. That newspapers suppress news which would reflect upon advertisers or influential citizens is the belief of 72.6 per cent; that newspapers publish avoidable inaccuracies is the belief of 50.5 per cent; and that they purposely falsify news is
the belief of 36.1 per cent.

Slightly more than half, 54.8 per cent, believe that newspapers give fair treatment to legislative news, 63.2 per cent entertain a like belief concerning religious news, and 74.2 per cent concerning labor news. Whatever the truth of these implied charges, it is evident that a significantly large proportion of those answering the questionnaire do believe the press guilty and that their faith in the integrity of the press is impaired to that extent.

Current practice in the treatment of anti-social news receives a degree of support somewhat surprising in view of the fact that in this field newspapers receive a great deal of criticism. Slightly more than half, 56.1 per cent, believe that newspapers publish too much sensational news. The fairly common practice of omitting the names of first offenders in minor crimes is endorsed by 63.2 per cent, and 66.2 per cent believe in the suppression of news the publication of which would harm the community.

That sex differences in attitudes toward questions of newspaper policy and practice were not great is shown by the fact that for only six of the twenty-five items is there a sex difference in response of more than 10 per cent. Somewhat greater conservatism is shown by the women than by the men through greater percentages of affirmative response to questions whether newspapers should suppress news harmful
to the community, whether newspapers should be licensed by the federal government, and whether publication of crime news leads to crime. On the other hand, the women, to a greater extent than the men, prefer larger headlines. They also favor more human interest material and more local news to a greater degree than do the men.

Seward and Silvers conducted a study in the belief in the accuracy of newspaper reports. Four brief reports of imaginary military or naval engagements were written up in four ways: favorable to United States, unfavorable, favorable to Japan, unfavorable. These were printed up as newspaper copy, pasted four different to a page in rotation, and presented to 209 women students in eight sections of economics and general psychology courses with instructions to rate them as to credibility from zero to one hundred.

In descending order of belief, the four versions ranked as follows: unfavorable to Washington, favorable, unfavorable to Tokio, favorable. Results were at least partially determined by three factors: (1) a tendency to believe one's own government; (2) a tendency to believe good news rather than bad; (3) a tendency to believe news adverse to its source.

---

Only in recent years have psychologists commenced seriously to study communications. Distinguished beginnings have been made in the areas of radio and cinema; but as yet the psychology of newspapers has received but little attention.

In an attempt to establish some principles, if confirmed they may become general laws, the noted psychologist Allport, and an assistant, concluded from the analysis of the Neutrality Act that gripping the attention of America in the fall of 1939, five tentative psychological principles.

The first principle is the issue becomes skeletonized. The extent to which simplification took place in the Boston papers which were studied was estimated as carefully as possible. Excepting the Christian Science Monitor, approximately eighty-six per cent of the space devoted to the Neutrality Act in the news columns dealt only with the repeal of the arms embargo. The conclusion is inescapable that in presenting issues of public policy newspapers simplify and skeletonize the situation to meet the reader's demand for sharpness and definiteness of structure. For explanation of skeletonizing, two related theories may be considered:

---

(1) Big-time news stories—so writes Charles Merz—are always stories of a fight (man against nature, man against man, or opinion against opinion). Fights are possible only when issues are clearly defined and when antagonisms are sharply drawn. Newspapers must dramatize and select in order to produce in their readers the emotional integration required for a good fight (or enjoyment of a good fight).

(2) The reader, like any organism, requires an economy of thought. William James spoke of the need of human beings for simply pathways to find their way through the tangled complexities of life. Clichés, stereotypes and fixed ideas are useful and necessary. Too many sides to an issue are confusing and fatiguing.

In the second principle the field of influence is well-structured. When a person sinks into his armchair and unfolds a newspaper, he is influenced not merely by isolated words, phrases, or articles; in a significant sense he is reading his paper. The copy of his paper in his hands becomes an indivisible and homogeneous stimulus-field to be "lived into" for half an hour after his evening meal. The impact of the paper upon him must, therefore, be expressed in terms of "pattern" or "configuration."

In principle three, Allport and Faden found that emotional restraint is greater among editors than among readers. It is a commonplace of newspaper psychology that an editor's decisions are governed by the caliber of his readers. But it is not frequently remarked that when an editor departs from his readers' standards, it is generally in the direction of greater restraint, caution, and judicial temper. The evidence for this assertion lies in a comparison of news and editorials with the Vox Populi. Where eighty-six per cent of the news columns and editorial space was devoted only to the issue of repeal, fully ninety-five per cent of the letters published dealt only with this question. Editors skeletonize; but readers do so to a still greater extent.\(^{13}\)

Principle IV - The intensity of interest varies in time. Thanks to the poll technique, studies in the change of opinion are increasing in numbers.\(^{14}\)

Principle V - Public opinion fatigues and presses toward closure. Public opinion, it has variously been said, is fickle, inert, and crowdish.\(^{15}\) On the basis of the authors' evidence they add one further epithet: public

---


\(^{14}\) Gallup became fully aware of this in the Truman-Dewey election.

opinion is fatigable. To them it seems likely that all these four attributes reflect one fundamental psychological tendency—the desire of the average man to be freed as quickly as possible from the tension of worry and annoyance in matters of public policy.

One of Ross' laws of social psychology comes to mind: the more intense a craze, the sharper will be the reaction against it. Applied to the Neutrality Act legislation, one might say that for the average man the more intense the feelings aroused by an issue involving public opinion the stronger will be the demand for closure. It is necessary to escape from tension as soon as practicable.

In an analysis of the press, reader habits, and reader interest, Bush and Teilhet roughly divided the content of the newspaper as follows: pictorial and graphical content, general news and information, comment, sports news, practical guidance, fiction, and advertising. They found that the average urban family reads about one and one half newspapers daily. This is in the ratio of almost three evening to two morning papers. With reference to the amount of time a reader spends in reading a daily newspaper it is not possible to say, for no objective technique of measurement has been devised. Average self-estimates reported in several studies

range from fifteen to forty-four minutes daily. It would probably be safe to assume that twenty minutes per daily newspaper is an approximate average, state the authors.

The reading interests of men and women with respect to different types of news may be roughly indicated in the following order: For men—war; weather, labor, sports, human interest; science, deaths, politics, crime; accidents; amusements (including the radio log), and business. For women: weather; human interest, deaths; accidents; amusements (including the radio log); crime, fashions, society and war.

Actually, the coefficient of correlation of men's and women's reader interests, when calculated on broad categories of newspaper content, is fairly high.

In a study made to establish tentative indices for the relation of newspaper reading patterns to the age, education and socio-economic status of readers, Schramm and White,17 with the caution that insofar as their data are representative, found that the following conclusions seem to be indicated:

1. In general, the amount of news reading tends to increase with age, with education, and with economic status. ...Education seems to make a greater difference in women's reading than in men's. Economic status seems to make a

---

greater difference in men's reading than in women's.

The importance of these three variables, age, education and socio-economic status, as well as sex, is further indicated by a study by Eysenck.\(^\text{18}\) He conducted experiments aimed at isolating, defining, and measuring primary social attitudes and finding their relationship to the aforementioned variables.

Nafziger, MacLean, and Engstrom\(^\text{19}\) made a study of the reading habits in a large city, a small city, and a rural community. The usual readership surveys supply some index of how much of what things is read by the potential reader. Generally, however, reader-interest indices are hard to interpret. This safe generalization can be made: this story apparently was better read than any other in the paper that day; these were the top-five pictures, or something of a similar nature.

Some studies provide scores for pages or sections of the newspaper and give figures for read-any-of-this or read-any-of-that. This indices are rough and (not so obviously) sometimes misleading. One very popular story in an otherwise non-popular group can increase greatly the percentages for the group of items.


Summaries of continuing studies present medians and ranges of percentage for various kinds of content. These summaries have the disadvantage of dealing with a heterogeneous, non-random sample of newspapers. Thus, a Thursday issue of one paper studied in 1940 is put on the scales with the Monday issue of another paper studied in 1947.

Figures by the authors show that much news goes unread by many potential readers. Popularity of local news was generally more noticeable in the small city and community weekly, however, than was in the metropolitan paper.

The country weekly published no comics and only one picture. Consequently only the small city and large city figures are given for features, comics, editorials and pictures. The popularity of pictures in the city paper is notable. Frequency distributions of attention given to comics in both papers showed, however, that a large portion of the readers read little or none of the comics, a few readers read a medium amount and a large part of the audience read nearly all of them. Intercorrelations showed that the reading of comics was essentially unrelated to other kinds of newspaper reading. This may mean, state the authors, that comics tend to hold a group of readers who might otherwise be uninterested in the newspaper.

Editorials in the large city paper did not attract extensive readership. The frequency distribution is L-shaped,
indicating a considerable group of readers who read no editorials at all. Very few read all of them.

The data in this study suggest that newspapers could profitably experiment with local pictorial editorials. Some problem which requires public action—drunken driving, for example—might be presented most effectively in an illustrated editorial.

On the average, only about a quarter of the feature material was read. Much of this reading matter came from national syndicates or the wire services; some of it, like sewing patterns and recipes, appealed to specialized interests.

No particular relationship is revealed between the amount of space given to a category and the extent of readership of that kind of news. Intensity of reading varied more in the rural paper than in the other two, a function perhaps of the relatively small amounts of space devoted to many of the categories.

News about education, crime or accidents, and politics was popular in all three papers, although these items were not sensational in nature. Even the accident story in the rural paper merely reported that a car driven by a woman had collided with another car at crossroads.

Under their section Who Reads What, some inferences regarding the popularity of various kinds of news can be
drawn from data on the kinds of people who read different kinds of news. Or, conversely, why do certain categories of news fail to reach a large proportion of the readers? Correlational studies show that reading of different kinds of news is interrelated, that a general newspaper reading factor operates, perhaps a generalized urge and ability to read, the authors believe. An example is shown in the relationship between the characteristics of readers which were significantly related to reading of all news.

Readers of Local News. One of the factors which appears to underlie reading of local news was an interest in public affairs. This observation is indicated by positive correlations of the city news which was read with registration for voting, ability to identify prominent members of the community, personal concern over the outcome of a foreign election, reading of magazines and the metropolitan daily, and a liking for serious talks and discussions on the radio.

Foreign News Readers. An examination of the readers of the small city paper which published twenty-eight column inches of foreign news revealed there were practically no differences among the various sub-groups who read news from foreign sources. Much of the same kind of undifferentiated reading of foreign news was done in the metropolitan area, except that the reading of foreign news was related positively
to magazine reading. Half of the avid readers in contrast to one quarter of the non-readers of magazines were above average readers of foreign news in the large city paper. There was no news from foreign sources in the country weekly.

Who Reads Comics. The weekly paper published no comics. The small city daily carried four comics strips and one cartoon. Comics reading was relatively unrelated to other types of reading in the small city paper, except in so far as their appeal tended to be general among all classes of readers.

Comics attracted particularly the middle income and occupational groups. This finding is in agreement with that of Schramm and White, both of which differ from a generalization by Gallup who stated that bankers, university presidents, professors, doctors, and lawyers read comic strips as avidly as truck drivers, waiters, and day laborers. The continuing study's findings are in agreement with the majority here.

The metropolitan Sunday papers were popular among comics readers. Small city comics readers were also avid radio listeners, followers of quiz programs, home-making,

---

20 Schramm and White, op. cit., p.157


22 Cited by Schramm and White, op. cit., p. 157
and popular music. Two pages, 156 column inches, of comics were carried by the metropolitan paper. This volume serves as a more stable basis for measuring the reading of comics than is true of the smaller papers, but few differences were evident among classifications of comics readers.

Who Reads Editorials. Editorials, in contrast to comics, are generally considered to be serious comments on important public affairs. Sometimes they are not. Reading of editorials in the small city paper was positively correlated with many variables, such as age, income, education, social activity, interest in world affairs, etc. The fact that so many variables were related to reading of editorials is important. Generally, they seem to indicate that the editorial reading group were more verbally skilled and more interested in serious public affairs than the group of readers who did not read editorials. None of these correlations was very high, however, most of them lower than .30.

Pictures were popular in all three papers. In two of the papers, the small and large city dailies, a considerable amount of space, 140 and 370 column inches respectively, was devoted to pictorial material. Woodburn23 states that size and subject matter are probably the most important factors which affect readership of newspaper pictures. Color is

---

another factor.

Nafziger and his associates conclude that comics reading appears to be the easiest to define and the surest to predict.\textsuperscript{24} The continuity involved in many of the comics, the fact that comics are generally located in the same position each day, and the "readability" of the comics may contribute to the stability of the following which the comics maintain. Little effort for immediate reward is a characteristic of comics reading.

It was shown in their study that the country weekly was more fully or intensely read than the small city daily and that the latter was more fully read than the large city daily. An increase in the size of the paper was accompanied, however, with a larger coverage of many kinds of news and feature stories and a greater appeal to various special interest groups among the readers than was true of the smaller papers. The most popular subject matter in the two smaller papers was the local news.

Depth of reading in the various kinds of news was relative to variations in socio-economic levels, media behavior, information levels, sex, and age. Certain categories like editorials, economic news, and politics, appealed particularly to the more serious reader who also tended to

\textsuperscript{24} Nafziger, MacLean, and Engstrom, op. cit. p. 539
be better informed, older, male, and a more consistent reader of other sources of information on public affairs. Society news, religion and news of leisure-time activities seemed to appeal mainly to women, older readers, ardent radio listeners, readers of special interest magazines, people with high school education, and readers who were relatively well informed on local affairs. Comics and pictures appealed to all groups.

In the Red Wing study, Nafziger and Barnhart conclude, among other things, that the pronounced popularity of local news—particularly war news with a local twist—is evident. It bears out the oft-made point that, for the reader, "significant" news is news that he can relate to his own life. And the more dramatic such news is, the greater will be his interest. Distant events, though they may be world shaking, must touch or be shown to touch the lives and experiences of the local readers before they get maximum attention among readers.

While the press is no longer looked upon as the "molder" of public opinion to the extent that it once was, the hypothesis remains that the content of the press exercises


an appreciable effect on human behavior. Construction of an index showing the visibility, coverage, and degree of favorability with which certain classes of industrial news are presented in the press at any given point makes it possible to test and refine this hypothesis by comparing fluctuations of the index with certain aspects of public behavior over a period of time, Harris and Lewis believe. The predictive value, the authors say, of this procedure with respect to some types of activity--travel by air, for instance--is appreciable. In a study of the foregoing, the authors obtained a co-variance of .5724. It is clear that these results, however startling they may appear to be, must be used with great caution, they warn the reader, explaining that seasonal variation in air travel may account for the co-variance, although

"this seems unlikely; and, indeed, seasonal variations in press reporting of air transport news may account for the co-variance, although this seems even less likely. These and other factors, including the limits of error of our own press estimates, must be considered and controlled before any flat assertion of causal relationship may justifiably be made from these data."

The co-variance, or Pearson product-moment coefficient

---


28 Italics inserted

of correlation, shows only the amount of similarity, in
direction and degree, of variations in corresponding pairs
of observations of two variables. The co-variance does
not measure causality. Another writer of considerable
repute as a researcher recently made this error.

English conducted a study to discover if readers of advertising in weekly newspapers react differently in their reading habits to home-set and foreign-set display advertising. The split-run technique was utilized as a simple "rider" experiment in connection with a reader-interest study of the Monticello, Iowa, Express. The hypothesis tested is that weekly newspaper reader-interest scores are unaffected by the typographical differences ordinarily existing between home-set advertisements and the prepared plates of national advertisers when both methods of type display local signatures.

English found that the reader interest scores for the four advertisements used were not affected significantly by the different typographic treatments.

The State University of Iowa School of Journalism, in cooperation with the Iowa Press Association, recently completed a readership survey of the Laurens Sun, an Iowa weekly newspaper. New light was cast on eight special

factors in weekly newspaper reading: when the paper is read, retention of the newspaper in the home, "possible" readers of the newspaper, subscription duration, overlapping circulation, what readers especially liked and disliked about the newspaper, depth of readership, and "most enjoyed" story on each page.

A total of 260 interviews (128 men and 132 women) were made in this survey. The sample was selected by stratified random methods from 96.1 per cent of the subscribers in the Laurens trade area. The interviewing method was that of "unassisted recall."

With reference to the depth of readership factor, for almost every editorial item (non-advertising or non-pictorial item), all or nearly all respondents who had read any part of the item claimed to have read from seventysix to one hundred per cent of it. Of all editorial items two jump stories, continued from page one to page 12, lost the most readers after the first quarter of the story had been read.

The results of an investigation of the degree of attention paid to various mass media and its possible effect upon the amount of information about current affairs which people have indicated -- reading in three written media--magazines, books, and newspapers--were related to variation.33

---

in information. Magazine reading appeared to be especially important. Relationships of radio listening and movie-going to information levels were not clear cut. In the groups studied, except that radio seemed to play an important role in information possessed by farmers.

Sex differences contributed substantially to variation in information. Men were particularly better informed than women on political figures. The socio-economic constellation of education, occupation, income and social participation was positively related to being informed.

Bigelow has developed the concept of absolute measurement, i.e., how many people are how much interested in what kind of feature? The word "absolute" is used by Bigelow in its strict sense of "self-existent," to emphasize the fact that this is not a comparative measurement.

Bigelow maintains that absolute-measurement of reader, competently developed, would enable the publisher...

"1. To obtain an absolute interest rating for any feature in print (including unpublished features of which he may pull proofs for research purposes);
2. To obtain this statement of interest for all the inhabitants of his market, for just his own circulation, or for any segment of either;
3. To make valid comparisons (within the limits of statistical tolerance) between ratings; and
4. To study in a single measurement the twin factors of numbers interested and intensity of interest."

Several new departures must be made, however, if this technique is to be of real service to newspaper management. First of all, research technicians must approach the study of reader-interest as a science in itself, rather than as a "little experiment" to check the validity of other types of audience measurement, as has heretofore been the case.

Before any useful findings may be expected, a careful program of pre-testing must be set up and the most effective field technique must be determined and adopted as standard procedure. Equal attention must be given the task of reporting to the management. Reports must be simple, easy to understand and use, and at the same time must not sacrifice any of the ratings' qualitative aspects. This procedure, too, should be standardized at the earliest moment. Pains should be taken to compute and report in every instance the statistical tolerance of the findings, and management should be cautioned against basing decisions upon inconclusive data.

Bigelow set up a table showing the differentiations between reader-interest study and readership study.

The reader-interest study:
...is technically known as a "self-rating attitude study;"
...is conducted with isolated, non-competing pieces of copy clipped from one or more newspapers or magazines;
...may be conducted with a cross-section of any desired "universe;"
...records the degree of interest expressed by each respondent in each item examined;
...is therefore a qualitative study;
produces data on "how many" are "how interested," which can be reduced to a single numerical interest rating, or which can (more properly) be expressed in a reader-interest curve;

produce an absolute rating, comparable with others developed by the same study or similar studies;

reports respondent's attitude;

is therefore most valuable in editorial research, since it is not affected by the eye-catching techniques which can secure a reader's attention despite his lack of any real interest.

The readership study:

is technically known as a "recognition-recall study;"

is conducted with a given issue of a single newspaper;

is conducted with persons having read some part of the issue being studied;

records which items in the issue each respondent recalls having read;

is therefore a quantitative study;

produces the standard readership rating, usually defined as "per cent of total readers who recalled having seen or read" any part of a given item;

produces a comparative rating, not strictly comparable with other ratings, since it is conditioned by the attention-value of adjacent items;

reports respondent's behavior

is therefore most valuable in advertising research, since it shows actual performance of advertisements under competitive conditions."

There are two basic uses to which absolute measurements of reader-interests can be put: (1) comparison of the interest expressed by the sample in various items studied, and (2) comparison of the interest expressed in the same item by various segments of the sample.

Studies of news readership, as they have come to be known and used, are mostly one-dimensional, says Schramm.35

They measure something akin to impressions; if a reader has sampled a story, he is counted as having read it.

Data--600 readership interviews, divided into three matched samples--200 on a weekly newspaper of a little less than 3,000 circulation, 200 on a daily newspaper of a little less than 10,000 circulation, and 200 on a large city daily with a circulation of over 300,000. To check the results of these interviews, he then made 450 additional interviews.

Interviewing technique was that introduced by George Gallup and developed by the Advertising Research Foundation in its Continuing Study, except as changes had to be made because of the purpose of this study. In order to insure a higher degree of validity, Schramm discounted a large number of stories because readers were not absolutely sure where they had stopped reading.

Schramm found that

1. A news story loses readers rapidly in the first few paragraphs. Thereafter, the curve of loss flattens out. If a reader gets past the lead and the first few additional facts, he is a good bet to finish the story.

2. As suggested above, the smaller a paper and the less its frequency up to once a week, the more likely it is to hold readers throughout a story. A weekly appears to hold its readers better than a daily, a small daily better
then a large daily.

3. In general, the longer a story, the smaller proportion of it likely to be read, and the faster it loses readers.

4. The average individual seems to read between a third and a tenth of the total news content of a paper.

5. High initial readership is no guarantee that readers will stay longer with the story or read a larger proportion of it.

6. Greater stylistic readability (as measured by the Flesch formula) seems to encourage greater depth of reading.

7. A feature style story seems to hold readers better than a straight news (inverted pyramid) type of story.

8. Skipping a story to another page is not so bad for readership as has commonly been supposed.

9. There is enough evidence in this study to warrant re-examining the use of subheads, bold-face paragraphs, and stars to break a story.

10. The formula $R \times D \times 100$ (in which $R$ is initial readership, $D$ is average depth of readership) seems to provide the most accurate index of news reading yet available, and suggests further study.

In another article, The Nature of News, Schramm discusses news as a part of the communications process.\textsuperscript{36}

and suggests that readers or listeners select news in expectation of a reward, which may be either the immediate "pleasure" reward of drive reduction or vicarious experience, or the delayed "reality" reward of "threat value." For any individual the boundaries of these two categories are not fixed and immutable, but that news of public affairs, economic matters, social problems, science, and education is generally read for delayed reward, and news of crime and corruption, accidents and disasters, sports and recreation, social events, and human interest for immediate reward.

Reading habits do appear to cluster around these categories. A person who is above average in his reading of one is likely to be below in the other. Reading for delayed reward seems to be a rather more sophisticated form of learned behavior which increases with education or similar experience and marks a development in the socialization of the reader. In either kind of reading, the ease of self-identification with the story is powerfully influential on the probability that a reader will select the story. And finally, that news is presented to its audience in the form of cues which index grey type or grey sound, and from which the audience selected those which seem to have the greatest predictive value in terms of reward.
Gallup presented a list of the means and the evidence then used by editors (1930) to judge reader-interest in feature and departmental material and indicated their unreliability.\footnote{37}

From the results from surveys conducted for six newspapers, three of them in cities over 500,000, two in a city of 100,000, and one in a city of less than 100,000 population, he presents about forty-six generalizations. Some have been disputed by other researchers, but it should be kept in mind that the findings of one particular study should not be applied without reservations to any paper or any individual, any more than a readership study of the Boston Globe should be applied without reservation to the Brookline Eagle or the Back Bay Ledger.

Gallup's study was the first systematic investigation of which items are read. The second dimension--how much of each item is read--was opened up by Schramm.\footnote{38} Both types of investigations have earned creditable reputations in the field of communications research. Having learned essentially what and how much people read, the complex problem of investigating the third dimension--why people read certain items--must now be faced. Through the efforts of a few

\footnote{37 George Gallup, \textit{op. cit.}, pp. 1-13}
\footnote{38 Supra, p. 112}
pioneers, commercial "motivational" research is slowly starting to develop, especially in the New York area. But, otherwise, motivational research has been well avoided by non-clinical investigators. This is especially true in journalism. The reason clearly seems to be that motivations, emotions, tensions, feelings and the like belong to that class of psychological phenomena which are not directly observable through the sensorium. They are elusive (and often illusive), and hence are not easily susceptible to a tabular method of analysis. Kay has investigated this dimension.39

The primary function of Part I of Kay's study is to set forth some basic psychological principles which are presumed to govern the news-reading process(es). These principles constitute the rationale for the formulation of the motivation-categories. Evidence in support of these hypotheses was culled from three major sources: (1) an original set of depth-interviews with adult residents of Palo Alto (Calif.) and surrounding area; (2) the only two prior reports discovered in professional literature on news-reading motivations; and (3) the application of the principles of psychology—chiefly Gestalt psychology—to the news-reading process.

No news is a discrete object, says Kay. That is, there is no such thing as a crime story or a public affairs item per se. A given news story may contain several elements, including crime, public affairs, accidents, health, etc. The elements may be explicitly mentioned or may be there just because of associations aroused in the reader's mind. It is only when an individual reader apperceives it to be a crime story that it is one.

Explanation of the term "apperception" would be in order here.

"We 'see' or become conscious of those elements which our personalities have prepared us or allowed us to 'see.' Further, we see not only what our experiences prepare us to see but also what we ourselves contribute to the item in order to make it intelligible. This happens because an item cannot stand by itself: it must be related in some way to other experiences we have had before it becomes meaningful. This process of perception-plus-intrepretation has been called apperception.40

In an attempt to modify Schramm's concept of immediate and delayed reward, Key proposes that a reward in news reading is obtainable only by learning something that was not previously known. He goes on to say that one can apperceive or anticipate something new or novel about the item but still fail to read it. Novelty and comprehension are necessary but not sufficient causes of news-reading.

40 Herbert Kay, ibid., p. 24
Regardless of whether a reader finds an item to be wholly and intimately related to his own life or merely fractionally related to his existence, if he understands the item—comprehends its significance—he has ego-involvement with the item. Ego-involvement is the generic term used here to indicate simply that the item has some meaning to the reader in the sense that it relates in some way to his history of experiences.

If that ego-involvement is very intimate, so that the news item bears directly and significantly upon the reader's own life, that species of ego-involvement relationship will be called personal-identification cathexis, Kay states. But when the ego-involvement is fractional and not very important to the welfare of the individual reader, it will be called entertainment cathexis.

Kay's study is an exploration into news-reading anxieties. His analysis of pre-tests strongly suggested that there would be little difficulty in applying the cathexis categories to responses from the main survey. Without hope of obtaining anything more than tentative data, his methods were devised for an exploratory study of news-reading anxieties and how they relate to cathexis-patterns of readership. The problem of anxiety was selected because it may well be the most important psychological factor governing news-reading.
behavior, and because apparently no one has attempted any kind of non-clinical empirical study of the problem. The experiment was considered to be highly successful in terms of four results: (1) the cathexis categories proved to be quite efficient; (2) the categories can be used for correlations with other variables; (3) some tentative but nonetheless highly provocative empirical data on news-reading anxieties were obtained; and (4) the great potentialities of projective techniques as disguised personality tests in communications research became apparent.

Kay found that the arousal of anxiety by a newspaper item does not always lead to an avoidance of that item.

Regardless of their educational level, higher anxiety persons read predominantly about themselves and their own problems—not about external things.41

The lower one's anxiety, the more he is concerned or preoccupied with the need to control problem situations in which he finds himself personally involved.

One difference between those with considerable anxiety who do not significantly repress and those with considerable anxiety who do repress, is that the former are more concerned with gaining control by accepting (reading) more items. These non-repressers also read more items of all kinds than

41 Walter C. Langer, Psychology and Human Living, 1943 (New York: Appleton-Century-Crofts, Inc.) p. 165
do the repressers. And the non-repressers are much more concerned with and read much more to gain control over situations than to extract entertainment from them. This is not hard to understand if we accept the notion that those who repress anxiety pretend that there is nothing to worry about, and hence will try to read not to overcome a threat which they don't see but rather for other reasons.

The very low anxiety persons, who also engage in little repression (because they have little to repress), read more items of all kinds than do either the repressers or non-repressers in the higher anxiety groups. But the higher anxiety non-repressers seem to read at least as much or perhaps slightly more to gain control than do the low anxiety non-repressers. This is understandable if we assume that they find more situations which need controlling (because they are threatening) than do the low anxiety readers.

The following conclusion was not obtainable through tabulation, says Kay, but nevertheless may be highly significant. It is that people who vitriolically complain about "sensational" news as being trashy, trivial, detrimental, ad nauseum, are usually those who suffer from a great deal of anxiety. This conclusion is supported by the fact that the low anxiety scorers were the ones who usually became
curious about or amused by the very items which the higher-anxiety scorers so vehemently rejected. It is quite true, of course, that not every objection to sensationalism is an anxiety-rationalization. Nonetheless, anxiety seems to be a remarkably pervasive determinant of news-reading behavior.

Nafziger, MacLean and Engstrom outline some of the useful tools for interpreting newspaper readership data.42 The aided recall method for studying newspaper readership has gained wide popularity since 1930 among newspaper publishers and advertisers. Underlying all these studies has been the assumption that the technique validly assesses how much of newspaper content the reader has read. Although investigations into the validity of the method are not in complete agreement, they generally tend to back up the assumption.

Split-sample reliability studies have shown that this method stands up well in relation to other measurements of behavior.

Analysis and interpretation are stumbling blocks in many present studies of newspaper audiences.

Two tools with which the authors are chiefly concerned are frequency distribution charts and factor analysis. The use of each requires that the newspaper content be broken down into meaningful categories.

Frequency distribution charts--It seems reasonable to infer that reading scores for newspaper content which distribute into a normal, bell-shaped curve differ sharply from material yielding U-shaped distributions.

The tool which seems to have the greatest promise is factor analysis. By observing intercorrelations of reading in the various categories, we can learn much about the basic organization of such reading. A factor is a generalization of a group of variables which are highly interrelated in the population under study. A certain amount of the variation of each variable in the group is explained by the variation of a single common factor. Factor analysis is simply a statistical method for defining a number of relatively specific variables in more general, more economical terms. This technique has been used extensively by psychologists--especially educational psychologists--in attempts to distinguish underlying differences in human abilities.

Plots of score frequencies have been most often used in connection with psychological testing. It will be helpful therefore to compare readership measurements with certain assumptions which underlie tests and test construction.43

The fact that reading a newspaper is typically a voluntary act differentiates it from the usual testing situation.

Most tests require the subject to make a response to each item. The newspaper reader, on the other hand, may have no other compulsion than a feeling that he should be informed, entertained, rewarded.

Coefficients of correlation—these are indices of intensity and direction of relationships. These coefficients range from -1.00 through 0.00 to 1.00. Given measurements along two variables (like reading of religious news and reading of comics), a perfect negative relationship (-1.00) means that every step-up in one measurement represents a step down in the other measurement. One of the basic assumptions underlying the correlation coefficient is that the relationship is rectilinear—that is, that the correlation between the two variables can truly be described in terms of a straight line. When two tests are highly correlated in a plus direction, they are measuring essentially the same thing. If they have a perfect positive correlation then it is assumed that one can be used interchangeably with the other.

Upon factor-analyzing their data the authors discovered that eleven logical categories could reasonably be considered as three constellations—so far as reading behavior was concerned. It supported too the hypothesis that the reader who tends to read much of one kind of news also tends to read

44, Supra, p. 107
much in many other categories of news. These analyses may give us a more useful psychological basis for describing content, especially in studies which include content, audience, and effect.

The authors' findings from two Minnesota studies shows that:

1. Histograms, or charts of frequency distributions, add considerably to one's knowledge of depth and character of reading in each kind of newspaper material.

2. The L-shaped histogram appeared to be most common—especially for reading in the metropolitan daily. This simply means that most people read very little of the material while a few people read a great deal of it.

3. The materials, as categorized, were sufficiently alike so that there were no significant negative relationships. Many were substantially positive.

4. Three factors proved adequate to account for the variation in readership in each newspaper. Tentatively, these might be described as:
   a. Easy-to-read entertainment material in the form of pictures and comics or the smaller papers and pictures, comics, and sports for the metropolitan newspaper.
   b. General and public affairs reading, with news about war-related activities, politics-government and economics dominating.
   c. Reading of news about people and society, arts and hobbies. For the small city audience, this third factor also contained reading of religious news; in the larger city, it included the reading of news about education.

5. The three factors were relatively unrelated in the small city; the second and third factor correlated substantially in the large city.
CHAPTER VI
READER-INTEREST SURVEY TECHNIQUES

The newspaper reader-interest study is a type of recognition test. It is carried out by presenting to the reader a fresh copy of yesterday's newspaper for recall-recognition of all items read. The maximum period of recall is, in the case of a morning newspaper, thirty hours, says Bush.1 There is abundant evidence that the responses represent a reliable measure of recall.

The author usually limits the size of his sample to approximately 500 because it is not possible to recruit and train more than forty-five or fifty field workers for an operation that must be completed in about eight hours. All tests for reliability, however, indicate that a sample of this size is adequate, states Bush.

To assure representative sex and age break-downs, a predetermined control of the sample must be exercised; for example, a disproportionate number of interviews must be made down-town in order to obtain the necessary men in the lower and middle age groups. (This procedure is to be preferred to interviewing men at home in the evening because it reduces the period of recall and eliminates confusion.

due to the effects of later editions and radio newscasts.)

Not only are certain physical factors important, such as presenting the newspaper where there is adequate light, but the psychological situation must be considered. That is to say, the field worker must endeavor to put the subject in the proper mental attitude for recall.

There is no evidence that fatigue is a derogatory factor in these interviews, although it has been proved an important derogatory factor in the advertisement "visibility" measurements of large magazines. ²

Reliability is chiefly dependent upon the skill employed by the interviewer in establishing the appropriate psychological setting for the interview. This is just another way of saying that the recruitment and training of field workers is the most important factor. Precise, fool-proof instructions must be prepared. Field workers should be given some preliminary practice in the presence of the supervisor so that erroneous assumptions and unsound procedures can be corrected.

It has been Bush's experience that senior and junior students of journalism and marketing make the best field workers when they are adequately paid for the work. They have an understanding of the purpose of the survey, know

something about the content of newspapers and appear to have respect for accuracy. When unpaid college students, however, are asked to do successive surveys, they appear to lose interest.

In another article by Bush\(^3\) he states that the sample used in recognition studies relating to daily newspapers is possibly as much an accidental as a random sample. This is because interviewing has to be done under difficult conditions. Since the survey should be made during the day following publication to avoid confusion caused by intervening new material, the period of interviewing is limited to a single day.

For this reason, and others previously mentioned,\(^4\) it is the usual practice to apportion the interviews by circulation districts, or routes, to control the sex quota, and, as far as possible, to control the age quota. Control of the sex quota requires either doing some of the interviews by a very accidental method in the business district or doing a certain proportion of interviews in the evening, the former practice being more general.

Bush’s experiment seems to indicate that a variation of this method would improve the sample for similar studies in large cities. The method would call for selecting the blocks

---

4 Supra, p. 126
and preparing the lists in the manner which he describes, omitting, however, the selection of every \( n \)th name on the cumulative list. Interviewers instead, would be permitted to interview any name on the cumulative list (i.e., any name on the division of the cumulative list that is assigned to the respective interviewer). A further improvement would call for random selection of blocks within the separate circulation districts and in proportion to the circulation within the respective districts.

To insure that quotes would be filled in those blocks in which a few interviewers might have difficulty, interviewers in rare instances could be permitted to cross the street to obtain an interview. Such an interview would be outside the selected block, but it is likely that the characteristics of a subscriber who lives across the street would match those of the subscriber included in the sample.

Bigelow\(^5\) explains how some commercial research agencies have broadened the scope of readership surveys and enhanced their usefulness. The suggested refinements are these:

1. That the readership study be conducted with a cross-section of the population living within the newspaper's city zone;
2. That a detailed report on the newspaper's penetration of its local market be made in connection with the readership study;
3. That the sample be broadened to include young people, fourteen to twenty years of age, and that their readership

---
be reported separately from that of adult men and women.

4. That the use of a pre-coded questionnaire be sub-
stituted for the practice of "marking up" an actual copy of
the newspaper on each interview;

5. That the readership of editorial and advertising
matter be reported by sex, age, and standard of living, and
be available on request by other controls; and

6. That a single, over-all readership rating be devel-
oped and reported for every item studied.

The refinements advocated above are based upon practices
now in use by commercial research firms and are in line with
the standards of accuracy and completeness demanded by ad-
vertisers. The research industry in the past decade has been
constantly revising its techniques, and media which espoused
research at a later date than newspapers have benefited from
the higher standards. There is no reason why newspapers
should not benefit similarly, helping themselves and their
advertisers alike, by refining the readership study. Efficient
marketing of newspapers' space and of advertisers' wares
cannot be achieved without efficient marketing research.
And improvement of editorial content through readership study,
it might be added, is similarly dependent upon better methods.

The technique of ascertaining reader preference by
having readers check opposite a list of features those
which they read "regularly," "occasionally," or "never,"
has been criticized as introducing errors based on subjective
response and imperfect conditions of recall. It has been
asserted, logically, that readers tend to overestimate the
degree of their "regular" readership. They do this partly
because of the difficulty in distinguishing between what is "regular" and what is "occasional" reading. It is believed, also, they tend to overestimate readership of features which they feel subjectively they "ought" to read regularly, such as editorials.6

Weigle7 contrasted the results obtained by the Gallup method with those returned in the mail questionnaire procedure, using the Arizona Republic of Phoenix as a control. The Arizona Republic mail questionnaire contained 102 items, of which forty-three were directly comparable with features listed in the Continuing Study survey. Since nine of the forty-three items were comics, they were combined and treated as a single feature, giving thirty-five features for which comparisons are available.

It was found that the correlation coefficient for men was .84 plus or minus a standard error of .03, and for women .80 plus or minus a standard error of .04, indicating that, on the whole, readers actually do read those items which they say they read "regularly." It is when one analyzes the specific items, however, that he discovers the operation of certain psychological factors that make the mail questionnaire method entirely unreliable, says Weigle.

6 The writer observed this tendency while participating in a 1600 readership survey of a Boston paper in November, 1953.

Both men and women heavily "overestimated" their readership in the mail survey by giving higher percentages of regular readership than they did in the subsequent Continuing Study.

Of particular significance to newspaper publishers is the manner in which respondents in the mail survey underestimated their regular readership of advertising.

There are some items in which the reader replying to the mail survey might be expected to give by rationalization a higher readership than he would give to the Continuing Study interviewer. Some of the features which the reader would feel he "ought to read," or "intends to read," or "Should keep up on," are editorials, health column, Biblical verse, Sentence Sermons, church notes, and market statistics. All of these have a significant "plus deviation" from the Continuing Study results.

Woodyatt\(^8\) also conducted a test comparison of the two techniques. The evidence of Woodyatt's study points toward the following chief conclusions: (1) that the nature of the material being tested may introduce a lack of candor, on the part of the respondents, which influences the results to a marked degree; (2) that regular readership percentages

---

disclose only part of a story which may, perhaps, have several parts; (3) that there are possibly better ways of determining readership habits than asking for a straight report of behavior; (4) that faulty recall does, as generally supposed, influence results markedly in a downward direction.

Woodyatt feels that low results for four of the five advertisements may have been caused by the respondents' embarrassment to admit they were interested in certain kinds of ads.

One of the ways in which to check on the inaccuracy of respondents is to ask them if they had seen material which they could not possibly have seen.

While there is some discussion of the results of his survey in terms of which type survey received a higher response in certain items, Woodyatt does not sufficiently verbalize his findings illustrated in his tables.

His conclusions are based upon trends indicated in his tables but they are inconsistent with the purpose of the study, i.e., a comparison of two techniques in readership research. The generalities, one to four listed above, are somewhat acceptable as generalities, but are not clearly substantiated by his findings.

Newspapermen have recently shown a tendency to divide

into two groups: those who are enthusiastic about public opinion research, and those to whom it is an anathema. This division has been sharpened by the 1948 election polls. Trescott gives from his experience some of the reasons both for the enthusiasm and for the distrust. Polls can be of great help to newspapers from a news, editorial, and public relations viewpoint; but at the same time their speed, accuracy, and cost must be brought into line with newspaper standards if they are to be the answer to the newspaperman's dream.

Trescott found, for example, that words which were commonly used by editorial writers were understood by only a limited number of readers. As a result, it has been desirable to explain briefly the meaning of such expressions as "tariff," subsidy," and "collective bargaining" whenever these subjects are treated.

One of the principal criticisms of present polling techniques, from a newspaper standpoint, is that it takes entirely too long to get the answers.

Some newspapers are deterred from polling by the expense. The author has found it costs his paper in the neighborhood of $70 a story, which is a lot more than many newspapers can stand.

English points out that the split-run copy method in

---

connection with a reader-interest survey may be used to test the effectiveness of a variety of news and advertising presentation techniques. Publishers who submit their newspapers to these studies should be aware that the split run method may be employed to reveal the relative effectiveness of publishing techniques without disturbing the reliability of traditional survey data.

A new method for determining newspaper audience consists in measuring the number of "impressions" of the specific media rather than the units purchased (circulation).

Like the Life studies which lead the field, this study done by Bush had for its aim the finding of the total number of advertising impressions to which a specific advertisement was exposed. The author feels that some newspapers may find it profitable to sell total advertising impressions rather than circulation—if advertisers are being educated to think in terms of "audience" as opposed to circulation.

The studies further suggest a more refined comparison of individual newspaper audiences. A newspaper with a large percentage of "home-read" circulation (read in the home whether delivered or purchased outside and brought in) may present a larger total "audience" than does a newspaper of


equal circulation whose large street vendor sales make for an unusually high number of casual readers (A.B.C. -- reported circulation that cannot be traced into the home either through delivery or by the news stand purchaser's bringing the copy home).

This can be demonstrated if we assume that each casually-read unit is read by a single person and if we multiply the total of home-read circulation by the average number of readers per copy and then add the two categories to obtain "audience," says Bush.
CHAPTER VII

Summary and Conclusions

Readability formulas attempt to measure the reading ease of an article. The one most adaptable for newspaper work, and the most controversial, is the one by Flesch. The formulas deal primarily with sentence length, word length in terms of prefixes and suffixes, and references to people. They omit the study of content, format, and organization, each of which is important to readability.

While editors have traditionally advised their writers and copyreaders to use "short words and sentences, they have not always required it, especially with foreign news.

Condensation--the art of telling the essential factors more clearly with fewer words--is ignored in readability formulas. It is employed to a considerable extent in some newspapers in England.

Editors and teachers have tended to carry readability scoring much further than its research has justified, while reporters' intuitive skepticism is subject to questioning.

Some researchers feel that professional newspaper executives do not possess sufficient criteria upon which to judge the effectiveness of stories in terms of reader comprehension. A few studies indicate the opposite.

In an attempt to correct the structural shortcomings of his earlier formula, Flesch developed a new readability
yardstick, composed of two formulas. Formula A is for predicting "reading ease;" formula B is for predicting "human interest." The reliability of the original and the simplified Flesch reading ease formulas was found to be quite high in a study by three authors. Reading ease with its components was analyzed quite reliably and human interest with its components was analyzed with less, though still fair reliability. Analysis of personal sentences showed the greatest lack of agreement between analysts.

Pursuing a lead discovered in his last revision, Flesch proposed to adapt the technique of readability measurement to the measurement of the level of abstraction, utilizing the fact that certain parts of speech are more frequent in abstract expression, while certain other parts of speech are more frequent in concrete expression. At this point of his research Flesch proposed that his measurement of the level of abstraction be applied to such areas as semantics, critical reading, literary appreciation, translation, rating of advertising copy, and propaganda analysis. In general, words related to abstractness are more "indefinite;" those related to concreteness are more "definite."

Other researchers devised tables for use with Flesch's level of abstraction, and criticized some of his assumptions. Flesch offered rebuttals to his critics.
Farr, Jenkins, and Paterson simplified the Flesch reading formulas and by so doing engendered a host of criticisms from Flesch and others. They counted the number of one syllable words per hundred words. In reply to attacks by Klare and Flesch, Farr and his associates conducted another experiment by which they established their original contention with specific reference to reliability of their simplified method. Employing sophisticated statistics Dunnette and Maloney factor-analyzed the original and the simplified Flesch reading ease formulas. They found that the counting of one syllable words could be done in about three-fourths the time required for counting syllables. They concluded that the new Farr, Jenkins, and Paterson simplified Flesch reading ease formula was truly simplified since it can be applied with a greater degree of accuracy and requires less counting time.

Swanson and Fox show that few validation studies using comprehension and retention as criteria have been made. They suggest that readability factors would have maximum effect when two or more positively related factors were varied. Easier words and shorter sentences, for example, should result in increases of comprehension, other things being equal. Motivational factors inherent in content, such as subject matter, probably are more important, generally, than
readability where individuals select what they want to read.

Readers judge text with the larger amounts of leading to be more legible and to be more pleasing for each length of line considered. Therefore, when a twelve-pica line width is used for eight-point type, one or two points of leading should be used if reader opinions are considered important, even though the leading does not increase readability.

For the second time in the present century, the physical format of daily newspapers is undergoing changes which in a large measure, may be attributed to the need to reduce production costs, especially expenditures for high-priced newsprint. It was found that newspaper readers have great difficulty in discriminating any difference between 7/8" and 1 1/16" shrinkage of pages when other page format variables are constant. The readers were more concerned that the newspaper be printed clearly so as to assure good contrast, and that the body type size not be reduced. The principal of ocular photography (known as the corneal reflection technique) is valid as a means of obtaining an accurate record of eye movements.

The function of the headline is to summarize in just a few words the most important elements of the story it covers. It is not the absolute size of the headline which is the determining factor as much as the position it has in the
immediate or temporal ground. Researchers have desired to measure headline display type from the point of view of "ease of seeing," but the validity of the techniques available for scientific appraisal remains unestablished among the principal experiments in the field. Some feel that the headline is the most potent factor entering into the formation and direction of public opinion in the United States. This is not established, however, by methodological experiments. One must not forget, of course, that the body of the news story itself is probably still the main source of opinion and interpretation. Recent studies indicate that too much importance has been attached to the role of headlines in promoting street sales.

Few schools attempt to teach correct newspaper reading habits to early and middle teenagers, the age when they first become "readers" of the newspapers rather than just the comics section. Five well-established points, what they call their principles, are stated by Allport and an assistant. They are: the issue becomes skeletonized, the field of influence is well-structured, emotional restraint is greater among editors than among readers, the intensity of interest varies in time, and public opinion fatigues and presses toward closure.

Nafziger and his associates conclude that comics reading
appears to be the easiest to define and the surest to pre-
dict. The continuity involved in many of the comics, the
fact that comics are generally located in the same position
each day, and the "readability" of the comics may contribute
to the stability of the following which the comics maintain.
Little effort for immediate reward is a characteristic of
comics reading. Certain categories like editorials,
economic news, and politics, appealed particularly to the
more serious reader who also tended to be better informed,
older, male, and a more consistent reader of other sources
of information on public affairs. Society news, religion,
and news of leisure-time activities seem to appeal mainly to
women, older readers, ardent radio listeners, readers of
special interest magazines, people with high school education,
and readers who were relatively well informed on local affairs.
Comics and pictures appealed to all groups. The pro-
nounced popularity of local news—particularly war news with
a local twist—is evident. It bears out the oft-made point
that, for the reader, "significant" news is news that he
can relate to his own life. And the more dramatic such news
is, the greater will be his interest. Distant events, though
they may be world shaking, must touch or be shown to touch
the lives and experiences of the local readers before they
get maximum attention among readers.

The concept of "absolute measurement" was developed by
Bigelow meaning, in its strict sense, "self-existent." There are two basic uses to which absolute measurements of reader-interests can be put: (1) comparison of the interest expressed by the sample in various items studied, and (2) comparison of the interest expressed in the same item by various segments of the population.

Schramm considers news as part of a communication and suggests that readers or listeners select news in expectation of a reward, which may be either the immediate "pleasure" reward of drive reduction or vicarious experience, or the delayed "reality reward of "threat value." Reading habits do appear to cluster around these categories. A person who is above average in his reading of one is likely to be below in the other.

Gallup's study was the first systematic investigation of which items are read. The second dimension--how much of each item is read--was expanded by Schramm. Why--the third dimension--people read certain items was investigated by Kay. He set forth some basic psychological principles which are presumed to govern the news-reading process. Kay maintains that it is only when an individual reader apperceives a story to be a crime story that it is one. Modifying Schramm's concept of immediate and delayed reward, Kay proposes that a reward in news reading is obtainable only by learning something that was not previously known. One can apperceive or
anticipate something new or novel about an item but still fail to read it. Novelty and comprehension are necessary but not sufficient causes of news-reading. Key's study is considered to be highly successful in terms of four results: (1) the cathexis categories he set up proved to be quite efficient; (2) the categories can be used for correlations with other variables; (3) some tentative but nonetheless highly provocative empirical data on news-reading anxieties were obtained; and (4) the great potentialities of projective techniques as disguised personality tests in communications research became apparent.

Three researchers outlined some tools for interpreting newspaper readership data. Two tools with which the authors are chiefly concerned are frequency distribution charts and factor analysis. Non-use or misuse of statistical tools have added confusion to a rather complex field.

Principles of representative sampling must be fully utilized in reader-interest surveys. The role of an effective interviewer is not to be overlooked. Some suggested refinements are: (1) that the readership study be conducted with a cross-section of the population living within the newspaper's city zone, (2) that a detailed report on the newspaper's penetration of its local market be made in connection with the readership study, (3) that the sample
be broadened to include young people, fourteen to twenty years of age, and that their readership be reported separately from that of adult men and women, (4) that the use of a pre-coded questionnaire be substituted for the practice of "marking up" an actual copy of the newspaper on each interview, (5) that the readership of editorial and advertising matter be reported by sex, age, and standard of living, and be available on request by other controls, and (6) that a single, over-all readership rating be developed and reported for every item studied.

The guiding principles for research and study in the field of journalism for measurement of readership and readability are no different from those which exist for research and study in any other field of psychological investigation. These are basically three in number: a firm adherence to the principles of experimental psychology (as exemplified by Kay), a firm adherence to the principles of sound statistical analysis (Ne fziger and others), and an adequate emphasis upon the theoretical framework within which any set of data is to find its meaning. Let us briefly consider each of these principles.

Too many investigators in the field of readership and readability contend or imply by their actions that the principles of sound experimental psychology need not be
applied to their field. To the extent they want to predict readership and readability for individuals, they are in the field of personality measurement, i.e., experimental psychology. One of the basic and fundamental rules in experimental psychology, as indeed it is in all science, is that we control all variables known or thought to have significance for the problem at hand. We can then permit alterations in one of the variables (the one we select for our predictor) and see what changes it causes in one or more of the remaining variables (our predictands; or, in more familiar terminology—dependent and independent variables).

It is realized, however, that it is quite difficult to control all the significant variables in journalism as it would be in the physical sciences. Some of the general rules to which an investigator or researcher should adhere are as follows:

1. The items which constitute the study should adequately measure what it is supposed to measure.

2. The men and women being studied should constitute representative or random samples of all men and women in that universe.

3. Sufficient cases should be studied to rule out chance variation as a principal component in the explanation of results.
4. Factors such as age should be held fixed or constant or taken appropriately into account so as not to obscure the significance of the results.

Principles of sound statistical analysis must be adhered to. Data collected through sloppy experimental procedures are absolutely worthless; and valid experimental data can be invalidated by inadequate or unsound statistical methods. Having once decided upon a sound experimental approach, it is well to consider the alternative types of statistical analysis that can be employed. Frequently it will happen that several experimental procedures, equally sound from the standpoint of control, may differ considerably in merit in terms of the type of statistical analysis that can be used in connection therewith.

It is the researcher's obligation, in the very beginning as well as throughout his study, to ask himself the following questions: How sound are the premises? How representative is the sample? Of what population is it alleged to be representative? How rigorous is the design of research? Do the conclusions necessarily follow from the design and the data? Over what range may the conclusions be generalized? What are the probable consequences of the conclusions? If the article is not an empirical study, does it offer hypotheses which are amenable to test? Can we move from
abstract variables to concrete data without too much logical violence? In other words, is he employing the Scientific Method?

Some researchers are fully scientifically oriented; other, not to a sufficient degree to ensure validity and reliability to their findings. If journalism research is to establish an accepted body of findings, all of its workers must be scientifically oriented in their methodology.
BIBLIOGRAPHY

A. BOOKS


Marks of a Readable Style, (Teachers College, Columbia University), 1943


Taylor, Howard B. and Jacob Scher, Copy Reading and News Editing, (New York: Prentice-Hall, Inc.), 1951, pp. 157-194


B. PERIODICAL ARTICLES


Farr, J. H., J. J. Jenkins, and D. G. Paterson, "Reply to Klare and Flesch re 'Simplification of Flesch Reading Ease Formula,'" Journal of Applied Psychology, 36:56-57, 1952


Fendrick, P. "Newspaper Reading Interests of High School and College Students," 34:522-530, 1941
Fisk, Marjorie, "Comparing Journalistic and Literary English," Journalism Quarterly, 10:202-208, 1933


"Reply to Criticism by Jenkins and Jones," Journal of Applied Psychology, 35:69, 1951

"Reply to 'Simplification of Flesch Reading Ease Formula,'" Journal of Applied Psychology, 36:54-55, 1952


Griffin, Philip F., "Reader Comprehension of News Stories: A Note on the Comment by Dr. Flesch," Journalism Quarterly, 28:497, 1951


Ludwig, Merritt C., "Hard Words and Human Interest; Their Effects on Readership," Journalism Quarterly, 26:167-171, 1949

Lyman, Howard B., "Flesch Count and Reading of Articles in a Midwestern Farm Paper," Journal of Applied Psychology, 33:78-80, 1949


Murphy, Donald R., "How Plain Talk Increases Readership 45% to 66%," Printers' Ink, 220:35-37, 1947


and Miles Tinker, "Effect of Line Width and Leading on Readability of Newspaper Type," Journalism Quarterly, 23:307, 1946


"Speed of Reading Nine-Point Type in Relation to Line Width and Leading," Journal of Applied Psychology, 33:81-82, 1949

"War Time Changes in Newspaper Printing Practice," Journalism Quarterly, 21:7-11, 1944


Ross, C. L., "Interests of Adults and High-School Pupils in Newspaper Reading," School and Society, 27:212-214, 1928


Woodburn, Bert W., "Reader Interest in Newspaper Pictures," Journalism Quarterly, 24:197-201, 1947