This is to certify that the

thesis entitled

A STUDY OF THE EFFECT OF REDUNDANCY

ON COMPREHENSIBILITY

presented by

Alice S. Horning

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in English

Paul E. Munsell

Major professor

Date August 30, 1977

A STUDY OF THE EFFECT OF REDUNDANCY ON COMPREHENSIBILITY

By

Alice S. Horning

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of English

1977

ABSTRACT

A STUDY OF THE EFFECT OF REDUNDANCY
ON COMPREHENSIBILITY

By

Alice S. Horning

In recent psycholinguistic research on reading, the concept of redundancy has been discussed in some detail. The research of Frank Smith and Kenneth Goodman establishes the importance of redundancy in the reading process. This study has focused on two aspects of the concept of redundancy: first, the development of a practical, operational definition of certain types of redundancy, and second, the investigation of the effect of artificially increased redundancy on the comprehensibility of texts in English for native speakers of English.

Redundancy was defined operationally on two linguistic dimensions, syntactic and semantic. Syntactic redundancy affects the structural predictability of texts, and can be increased by revising sentences into the more predictable noun phrase-verb phrase order and by, in general, replacing in the surface structure content which is often omitted by optional transformations. Semantic redundancy involves the repetition of key ideas through several types of clarifying or defining words or phrases.

One major hypothesis and two corollary hypotheses were investigated. The major hypothesis is that increasing syntactic and semantic redundancy increases the comprehensibility of text. The two corollaries are, first, that increased redundancy increases

reader interest in a text, and second, that increased redundancy improves the reader's evaluations of writing quality in a text.

Three texts of approximately three hundred words on three different topics were rewritten in four versions: an original version, a version with syntactic redundancy alone increased, a version with semantic redundancy alone increased, and a version with both syntactic and semantic redundancy increased. All passages were controlled for total length, sentence length, and vocabulary level. Each passage was prepared as a Cloze test in which fifty nouns and main verbs were deleted. The subjects were 240 freshmen at Wayne State University enrolled in Freshman Composition during the winter term, 1977. All subjects were native speakers of English.

The Cloze test scores were analyzed using the analysis of variance and the results of the interest and writing quality ratings were analyzed using regression. An effort was made to correlate background data (age, sex, major, and so on) on subjects to Cloze scores, using regression and the correlation ratio.

The major finding of the study is that increased redundancy seems to increase comprehensibility in a limited and qualified sense. Increased redundancy of both types significantly improved comprehensibility on two of the three passages studied and made no significant difference in the third case. Neither of the corollary hypotheses was confirmed by the data collected. An effort was made to explain the lack of significant difference in the third passage, but no statistically defensible explanation could be derived. The raw data seem to suggest that the redundancy created by the reader's prior knowledge of the topic and interest in it controls the effect

linguistic redundancy has on a text. These factors require further study. The operational definition of certain types of redundancy set up for this investigation could serve as a model for defining other types of redundancy and provide a basis for further research.

DEDICATION

The achievement that this volume represents can be attributed to the efforts of four people: to Julia Falk and Howard Helsinger for requiring me to seek the limits of my abilities, and to Paul Munsell and George Bland for always believing I would be successful in my search.

ACKNOWLEDGMENTS

No dissertation is ever completed without the advice and support of a number of people. In this case, the aid of the following people is especially acknowledged: to Dr. William Landrum and Ms. Becky Wiggins of the Wayne State University Computer Center, for their help with the computer; to Ms. Barbara Reetz for her assistance in scoring the data; to my colleagues in the English Department at Wayne State University, for their assistance in collecting the data and their general support of and interest in my work; to Mr. William Horning of Hastings, Michigan, for teaching me essential mathematics and statistics; to Ms. Margaret Maday for typing the manuscript; and to the following individuals for support above and beyond the call of duty or friendship: Jeffry Aronsson, Charles Baxter, Michael Bell, Barbara Reetz, and Roy Weitzel.

Finally, acknowledgment is made to my husband, Arthur Horning, for his continued willingness to sacrifice his own happiness to my degree, and for his perfect confidence in me at all times.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

STATEMENT OF PROBLEM, REVIEW OF LITERATURE

1. Introduction

At its inception, the goal of this project was to try to provide suitable beginning reading materials for adult students of English as a second language (ESL). It was thought that increasing the redundancy of sophisticated materials might make them more accessible to ESL readers. Two underlying problems became clear: the fact that an operational definition of redundancy was not available and the fact that little was known about the effect of redundancy on English text for English speakers. In response to these problems, the study had two major goals: first, to develop the operational definition of redundancy that was lacking, and second, to investigate with several hypotheses the impact of redundancy on the comprehensibility, interest and stylistic quality of a text in English for adult native speakers of English.

The subjects in the main study were 240 members of the freshman class enrolled in Freshman Composition at Wayne State University in the winter term, 1977. Three passages of text on three different topics were prepared in four versions each: an original version, a version with syntactic redundancy increased, a version with semantic redundancy increased and a version with both syntactic and semantic redundancy increased. Redundancy was operationally defined as having

a syntactic and semantic component. Syntactic redundancy was defined as the structural predictability of sentences in the text, and specific guidelines were developed to increase this type of redundancy. Semantic redundancy was defined as the repetition of information in a passage by the addition of clarifying phrases of several types. The types of clarifying phrases which could be used to increase semantic redundancy were defined.

Each subject in the study received one of the twelve passages at random. The passages were controlled for length, readability level, and number of items deleted. The exact items deleted in each version were virtually the same. Each passage was prepared as a Cloze test using a one-in-six deletion formula and choosing for deletion nouns and main verbs. The subjects had twenty minutes to read the passage and fill in as many blanks as they could. Subjects were also asked outside of the timed session to provide information about themselves (age, sex, major, etc.) and to answer, after the reading, two rating questions on their interest in, and evaluation of, the writing quality of the passage they read.

The data thus generated were analyzed using three statistical procedures: analysis of variance on the Cloze scores and regression and the correlation ratio on the background information on subjects with respect to their Cloze scores. The results of the analysis show that the operational definition of redundancy used in the study is a viable definition and that redundancy, as defined, seems to improve the comprehensibility of English text for native speakers of English. That is, the major hypothesis of the study, that increasing redundancy increases comprehensibility, is confirmed by the data, albeit in a

limited and qualified sense. The corollary hypotheses that increasing redundancy increases interest and subjective evaluation of the writing quality of the text are not confirmed by the data.

The study relies heavily on previous work done in reading and information theory. Although only a small number of scholars have written extensively about the role of redundancy in reading, their descriptions of redundancy provide essential background. The work of these scholars is reviewed in detail in chapter one, pursuant to a more complete statement of the problem under study here. Chapter two describes the methodology used in the formal study, together with background on the development of the procedures used through pilot studies. The third chapter provides the results of the formal study, and chapter four is an analysis and discussion of the results and the conclusions to be drawn from them.

2. Statement of Problem

In the recent psycholinguistic research on the reading process to be reviewed in the next several pages, the concept of redundancy appears again and again. One of the important practical questions to be raised about this concept is, can redundancy be used to improve reading as the process of getting meaning from print?

A prior and more fundamental question appears to underlie these questions and others: what precisely is meant by the term "redundancy" as it is being used here? The fundamental nature of redundancy is quite perplexing. The extant definitions are interesting and useful as descriptions, but are difficult to apply in the present practical context. Even if a workable, practical definition

can be developed, little work has been done to investigate the use of redundancy to facilitate reading. Moreover, because of what is already known about redundancy, the research also concerns itself with what the reader brings to the text, and with the readers' responses to the text in terms of their interest in it and their judgment of its quality. These reactions are crucial to the central question of increasing comprehensibility.

Thus, in addition to formulating a practical definition of redundancy, the research also deals with the question of the effect of increased redundancy on comprehensibility of English texts for adult native speakers of English. Two fundamental definitions will be constructed to serve as a base for the major hypothesis of this study: first, a definition of redundancy in practical, operational terms, and second, a definition of the term "comprehensibility," also in practical, operational terms. The research reported here investigates the following three hypotheses: increased redundancy yields increased comprehensibility; increased redundancy yields increased interest in a text; and increased redundancy improves the subjective evaluation of the quality of writing in a text.

3. Review of the Literature: Definitions of Redundancy

Extant definitions of redundancy come from two major fields: psycholinguistics and information theory, and from a limited number of researchers. The most extensive discussions of redundancy as it relates to language in general and to reading in particular are provided by Frank Smith, Kenneth Goodman, Colin Cherry and Wendell Garner. The work of these researchers along with one or two others

is reviewed here and an operational definition of redundancy is presented.

Redundancy is a key factor in the psycholinguistic theories of reading developed by Frank Smith, Kenneth Goodman, and others. One reason that Frank Smith's work is relevant to the research at hand is that his theory of reading relies heavily on the concept of redundancy. A skilled or fluent reader is defined by Smith as " . . . one who makes maximum use of redundancy . . . "[1] to get meaning from print. Having provided this definition, Smith goes on to devote much of his discussion in Understanding Reading to defining and detailing the nature of redundancy. In addition to a definition and explanation of redundancy, Smith provides an explanation of what a fluent reader seems to do with the redundancy of language, and he makes an interesting remark about the unconscious nature of a reader's use of redundancy.

Smith's definition of redundancy is briefly but clearly stated early in Understanding Reading. "Redundancy exists," he says, "whenever information is duplicated by more than one source."[2] This is a very straightforward, but also a very general statement. The implications of the definition in everyday life are quite obvious--any form of repetition is a simple kind of redundancy. In reading, however, the nature of redundancy is somewhat more complex, since there are, according to Smith, at least four alternate sources of information, and these sources can lead to redundancy in printed material.[3] These four sources of redundancy are: visual information, or actual text displayed on the page; orthographic information, or what the reader knows about letter sequences or spelling; syntactic

information, or the information which can be obtained through sentence structure; and semantic information, or that which can be obtained from context. "To some extent," Smith notes, these sources "provide overlapping information"[4] and are, therefore, redundant.

A fluent reader uses redundancy to get meaning from print. Indeed, according to Smith, people cannot be fluent readers unless they can use redundancy to get meaning:

> It can be shown that fluent readers make use of all the different aspects of redundancy because they require less visual information to identify letters in words than letters in isolation, and less visual information to identify words in meaningful sequences than in unrelated sequences of words. It can also be shown that fluent readers are capable of immediate word and meaning identification. Immediate word and meaning identification are not possible unless the reader is able to make use of orthographic, syntactic and semantic redundancy.[5]

But, this use of redundancy by the reader is unconscious:

> . . . it is not suggested that a reader is <u>aware</u> of his knowledge of sequential redundancy, any more than he is aware of the decision-making process that is involved in reading or any other form of perception. But . . . the fluent reader must indeed be regarded as possessing such a knowledge . . .[6]

Thus, readers are apparently unaware of their knowledge and use of redundancy in the process of getting meaning from print.

From Smith's work, it appears that redundancy is to some extent language ability specific (i.e., tied to one's unconscious knowledge of a language and ability to use it), that the reader's use of it is unconscious, and that in reading, there are at least four sources of it.

But there is more to the psycholinguistic view of redundancy than this. Kenneth Goodman's theory of reading, like Smith's, rests in large part on the assumption that the reader makes use of the redundancy inherent in printed text. Although Goodman's view overlaps that of Smith to a large extent, his remarks provide further insight into the nature of redundancy.

Goodman provides a more detailed account than does Smith of the sources of information in printed language and how the reader brings them to bear on the process of getting meaning from print. From this process, as Goodman views it, a definition of redundancy and a full understanding of its impact on reading are clear. Goodman also comments on the fact that a reader's ability to use redundancy is largely based on unconscious language abilities.

In general, Goodman's view of reading is aptly summed up in the title of one of his articles: reading is a "psycholinguistic guessing game"[7] where the guesses are based on the interaction of thought and language. The proficient reader supplies a great deal, in Goodman's view, and requires very little from the visual display on the page. There are, again, four sources of information (cue systems) derived from the interaction of thought and language. These are Goodman's four cue systems:

> There are really four kinds of cue systems that operate in reading to cue meaning. These are (1) cue systems within words, (2) cue systems in the flow of language, (3) cue systems within the reader, and (4) cue systems external to language and the reader.[8]

Goodman's first cue system incorporates Smith's first two sources of information, and Goodman's last system is not included by

Smith. However, the substance of these two views is the same in that both Smith and Goodman feel that it is these redundant qualities of language which enable the proficient reader to get meaning from print at high speed, and in that they agree that two of the major sources of redundancy are syntax and semantics.

Goodman's definition of redundancy grows out of his analysis of the four cue systems and it is somewhat more precise than Smith's. Goodman also explains the impact of redundancy on reading:

> Communications theorists use the word redundancy in a special sense to describe a tendency of languages to restrict the sequences in which language symbols can occur, to provide several cues to the same bit of information, and thus to be less than 100 percent efficient in the amount of information transmitted per unit of language. . . . This inefficiency or redundancy has two important effects on reading. First, it provides the reader with the repetitious cues we noted earlier. . . . Second, redundancy provides a narrowing of elements in the language that can fill certain slots.[9]

Again, use of inherent redundancy is essential to fluent or proficient reading for Goodman's theory just as it is for Smith. Indeed, Goodman is again in agreement with Smith that the use of redundancy to get meaning from print, to predict, to play the psycholinguistic guessing game, is unconscious:

> At any point in time, of course, the reader has available to him and brings to his reading the sum total of his experience and his language and thought development. This self-evident fact needs to be stated because what appears to be intuitive in any guessing is actually the result of knowledge so well learned that the process of its application requires little conscious effort. Most language use has reached this automatic intuitive level. Most of us are quite unable to describe the use we make of grammar in encoding and decoding speech, yet all language users demonstrate a high degree

> of skill and mastery over the syntax of language even in our humblest and most informal uses of speech.[10] [Emphasis mine.]

Goodman is saying here that readers are able to make use of redundancy because of their unconscious knowledge of their language (i.e. linguistic competence) and because of their conceptual and experiential background. Because Smith and Goodman are in agreement on this point, it is important to consider all that the reader brings to a text when the effect of redundancy is examined.

The psycholinguists' definition of redundancy can be summarized by the following points: redundancy exists whenever information is duplicated by more than one source, and in the case of reading, information is duplicated by at least four sources or cue systems. Redundancy, or the ability to use it, is crucial to proficient reading, but is, at the same time, a largely unconscious matter. It appears that some aspects of redundancy--sequential letter and word constraints, in particular--are language specific. Finally, it appears that the ability to use the redundancy of print is closely related to linguistic competence.

This consensus among the psycholinguists is helpful insofar as it provides us with a clear, albeit theoretical, description of the key characteristics of redundancy. To investigate the concept of redundancy further, it must be moved out of the realm of theory and applied to the practical problems discussed above.

A second set of definitions of redundancy comes from information theory. Colin Cherry provides two initial words of caution concerning the definitions of redundancy from information theory. First, he notes the quantitative treatment of redundancy in

information theory:

> In communication theory, redundancy is treated mathematically, the syntax being described, not necessarily as a linguist would commonly view it, but as a set of conditional probabilities.[11]

Then, he points out the difficulty of making any generalizations about redundancy:

> The relationship between the whole structure of a language (the morphemic, syntactic, grammatical formalism) and the outside world associations (its semantic functioning) is extremely complicated; it is essentially empirical and above all, varies between different languages. Again, redundancy is built into the structural forms of different languages in diverse ways.[12]

With these cautions in mind, when we set about to look at the definitions given by Cherry, and other information theorists, we find a substantial consensus. Cherry says:

> Simple repetition of a signal is the most elementary way of introducing redundancy . . . Briefly, redundancy is a property of languages, codes, and sign systems which arises from a superfluity of rules and which facilitates communication in spite of all the factors of uncertainty acting against it.[13]

J. R. Pierce says:

> English text, and most other information sources are redundant in that the messages they produce give many clues to the recipient. A few errors caused by replacing one letter by another don't destroy the message because we can infer it from other letters which are transmitted correctly.[14]

and Wendell Garner gives perhaps the most complete and complex definition of redundancy in his attempt to apply concepts from information theory of psychology. In Uncertainty and Structure as Psychological Concepts, Garner says:

> Total redundancy is rather simply defined. It is simply the difference between the total

> possible uncertainty of a set of variables and the actual uncertainty. It is convenient, however, . . . to distinguish between distributional constraint and the constraint between variables, and in the present context this latter form will be called sequential constraint . . . .[15]

Frank Smith provides an explanation of Garner's distinction between distributional and sequential constraint or redundancy, because Smith also uses the distinction. On distributional redundancy, Smith says: "Distributional redundancy is associated with the relative number of times each of the alternatives that constitute the uncertainty of a particular situation can occur."[16] In language, the alternatives can be letters or words.[17] On sequential redundancy, Smith says: " . . . sequential [information is] our knowledge of the way words are constructed . . . the sequential redundancy of English words is enormous."[18] Thus, distributional redundancy deals with how often a particular letter or word may occur in print, and sequential redundancy deals with the information we have about what letters or words can follow a particular letter or word. Garner, like Cherry and Pierce, feels that redundancy is duplication of information, although Garner's explanation is considerably more technical and mathematically precise.

This consensus among information theorists is certainly convenient and helpful insofar as these scholars have used a number of different kinds of statistical analyses to back up their definitions,[19] and insofar as there is also a partial consensus among information theorists and psycholinguists. All the views considered thus far seem to share these common points about redundancy: it exists in printed language; it varies somewhat from

language to language; and a reader or receiver of a message uses redundancy to get the message or meaning correctly, but this use of redundancy is unconscious.

At a number of points, the information theorists state that they have no answers to some very important questions. Garner notes, for example, that sequential redundancy in English is about 50% and total redundancy is about 60%,[20] but that these figures are only estimates. Garner believes there is no way to get more accurate, relevant figures, since we know no way of measuring the redundancy of printed English directly.[21] Garner notes that, even in dealing with letter and word sequences, only estimates are possible:

> The direct measurement of redundancy of printed English is, of course, essentially impossible. The number of possible different letter sequences, even as short as ten letters, is prohibitively large for any direct count of them. It has therefore been necessary to use indirect techniques for estimating redundancy of language.[22]

It is important to note that estimates have been made of letter and word redundancy. Shannon's "guessing game" technique, reported by Garner,[23] gives a lower bound estimate for letter redundancy. Other studies replicate Shannon's work,[24] and give a general estimate of sequential redundancy at 50% for printed English. A second technique, called multi-variate analysis, was tried by Garner and his colleagues.[25] But again, these figures are only estimates.

Moreover, Garner points out in his critique of the guessing game technique that

> how close the lower-bound estimate comes to the true redundancy of the language is obviously a <u>function of the skill of the guesser in using the redundancy of the language</u>.[27] [Emphasis mine.]

Thus, these statistical estimates do not account for the reader's skill in making use of what is in the text. Frank Smith is in perfect agreement with Garner on this point, as well as on the general problem of measuring redundancy:

> . . . I do not know of anything that is being done on the measurement of redundancy . . . . Although technically redundancy can exist wholly within a piece of text, I think what is of more importance in reading is the use the reader is able to make of the redundancy. Put another way, there is a redundancy that exists between the text and what the reader knows already. And it is impossible to measure or even consider this type of redundancy without considering what the individual reader brings to the task.[28]

Present measures of redundancy do not deal with what the reader brings to the task in a direct way.

Garner also says that no experiments have been done in information theory or psychology to test the effect on learning that adding or reducing redundancy has.[29] Thus, no research yet tells us whether increasing redundancy will improve learning, readability of a passage, or reading ability.

One important point Garner makes is that there is higher redundancy in some language samples than others. He cites a study in which the language of air traffic controllers was analyzed and found to have about 96% redundancy.[30] This is encouraging for two reasons: first, it means that a plane is a lot less likely to crash because of communications problems between the air traffic controllers and the pilots than we had perhaps thought. More to the point, however, it suggests that increased redundancy effects better communications for air traffic controllers. Thus, redundancy might be artificially increased in a text, resulting in an improvement

in communication between a writer and a reader.

This suggestion about mechanically increasing redundancy is only theoretical at this point. The problem at this juncture lies in the difference between theory and practice. The descriptions by Smith and Goodman, and the discussion of redundancy in information theory provide theoretical, descriptive background information about redundancy. None of these scholars has attempted to apply the concept of redundancy to a reading problem in the specific practical terms which are the focus of the present research. Moreover, Garner and Smith, at least, indicate that what the reader brings to a reading task is essential, yet they cannot describe or measure it with respect to redundancy and its effects. By taking this background material and using it as the foundation for an operational definition of redundancy, it will be possible to discover whether or not redundancy can be used to increase the communication between a writer and a reader.

The syntactic and semantic systems are probably the two sources of information in the text that fluent readers rely on very heavily in the process of getting meaning from print. These two sources of redundancy are the focus of the operational definition to be used here. The key ideas underlying the notion of redundancy in this operational sense are predictability and repetition. Predictability tends to operate more importantly in syntactic redundancy--in syntactic terms, redundancy might be increased by making sentence structures conform to the more predictable NP-VP pattern, and by otherwise increasing the predictability of the grammatical structure.

The relationship of predictable syntactic structure to comprehension was established in an early study done by Robert Ruddell.[31] In his study, Ruddell examined the effect of increased structural redundancy on the reading comprehension of fourth graders. By comparing Cloze scores on two passages, on the same topic but with differing levels of structural redundancy, Ruddell found significantly higher comprehension on the passage with greater structural redundancy. Structural redundancy and its effect on comprehension are described by Ruddell as follows:

> . . . the occurrence of one structural element delimits the range of elements following it. For example, the occurrence of a noun group will be followed with greater probability by a verb group than by another noun group . . . . One might expect reading material possessing structural elements which occur with higher frequency and with greater sequential constraint to result in greater redundancy and thus be more easily comprehended than reading material possessing a lower degree of structural redundancy resulting from elements occurring with low frequency and less constraint.[32]

Ruddell's work makes clear that increased syntactic redundancy improves comprehension for children, and this study looks for the same results for adults. The specific guidelines followed in rewriting texts to increase the syntactic redundancy are discussed in chapter 2.

By contrast to predictability, repetition tends to figure more importantly in semantic redundancy. In semantic terms, redundancy might be increased by adding phrases which provide examples, specifications, or clarifications to the text. These additions are intended to repeat information by providing the same ideas in several forms, thereby making the meaning more accessible. The specific

guidelines used in rewriting texts to increase semantic redundancy are discussed in the methodology section.

The concepts of predictability and repetition overlap to some extent, particularly when both methods of increasing redundancy are applied to a passage at the same time, as is the case in the present study. In operational terms then, redundancy is defined as syntactic modification of a text to increase its predictability and semantic modification of a text to increase its repetitiveness. The application of this definition to a group of texts should indicate whether or not increasing redundancy in these ways improves the communication between a writer and a reader.

4. Review of the Literature: Definitions of Readability

The amount of communication between a writer and a reader is often referred to as the readability of a text. To look at readability in another way, it is classically defined as the relative ease or difficulty of a text for a particular reader. The classic definition, which is illustrated below, has evolved into a superficial concern with the ability to predict the suitability of a text for a particular group of readers. For this reason, the accessibility of the message of a text will be referred to as comprehensibility instead of readability. Thus, comprehensibility encompasses both the classical concept of readability and true comprehension, i.e. whether the reader gets the message of the text. In studying redundancy, the focus is more on comprehension and less on predictions of suitability of text. However, the key issues in readability have been controlled for in the research.

One of the classic definitions of readability comes from the work of Jeanne Chall and Edgar Dale, reported in Chall's monograph on readability. It is an admittedly broad definition:

> In the broadest sense, readability is the sum total (including the interactions) of all those elements within a given piece of printed material that affects the success a group of readers have with it. The success is the extent to which they understand it, read it at an optimum speed, and find it interesting.[33]

Despite its breadth, there are two important features of this definition which make it relevant to the points at issue here. First, Dale and Chall are concerned not only with the elements in a text but also with the way the reader interrelates those elements. Second, Dale and Chall's definition of success in reading indicates that readers must find the content of the text interesting in order to be successful with it.

Other scholars, concerned with readability, have expanded the definition to make clear the nature of various components or "elements" of a readable text. John Bormuth expands the definition of readability in his discussion of the aim of the early research, which was, principally, to find "formulas which educators could use to determine if materials were suitable to their students."[34] In a way, of course, Bormuth is making essentially the same point as Dale and Chall. That is, Bormuth is saying that readability in part involves the appropriateness of the material for the students who will read it.

For an analysis of the various readability formulas available in 1968, Bormuth provides the following list of factors affecting readability: vocabulary complexity, word length, morphological

complexity, Latin base syllables, abstractness, frequency of word occurrence, grammatical complexity, syntactic depth, modifier distance, transformational complexity, and contextual variables.[35] Reading Research Quarterly contains an article by George Klare, who summarizes the various readability formulas now in use and their merits and flaws.[36]

Two of the most widely used formulas for measuring readability are the Flesch and Dale-Chall formulas, developed in 1943[37] and 1948[38] respectively. Flesch's formula has appeared in at least two versions, the second of which appeared in the same year as Dale and Chall's. Flesch's revised formula measured Reading Ease through a count of the number of syllables per hundred words and a count of the average number of words per sentence. Flesch also had a measure of Human Interest which was calculated by taking counts of the number of personal words[39] per hundred words and the number of personal sentences[40] per hundred sentences. The Reading Ease portion of the Flesch formula was one of the most widely used readability formulas.

The Dale-Chall formula was also a revision of the first form of Flesch's formula.[41] Like Flesch's own revision, Dale and Chall counted sentence length. They also use a Dale score, which was a percentage of the number of words in a passage outside of Dale's list of three thousand familiar words.[42] Also like Flesch's revised formula, the Dale-Chall formula was widely used to measure readability. A number of revisions and recalculations of these formulas have been carried out since 1948. In addition, a number of new formulas for readability have been developed, and are reviewed by George Klare.[43]

Klare comes to an interesting conclusion after his lengthy report. He finds that a formula for readability which incorporates the two variables of word length and sentence length is quite good for making predictions about readability. In view of the number of factors listed by Bormuth, this may seem like a great oversimplification. But Klare responds as follows:

> It may seem surprising that counts of the 2 simple variables of word length and sentence length are sufficient to make relatively good predictions of readability. No argument that they cause ease or difficulty is intended; they are merely good indices of difficulty.[44]

Klare is suggesting that the elements considered by the Dale-Chall formula can predict the ease or difficulty of a text for a group of readers. However, the research of Smith, Goodman and others suggests that ease or difficulty of text is in fact a function of other factors. One of these other factors involves what is in the text itself or what might be called the inherent or linguistic redundancy of the text. A second factor affecting the relative difficulty of the text for a reader or readers involves what the reader brings in terms of experiential as well as linguistic sophistication. This second factor might be called a kind of pragmatic or experiential redundancy. What the research presented here suggests is that ease or difficulty of the process of getting meaning from print seems to be controlled substantively by both linguistic redundancy and the experience and background the reader brings to the text.

Despite the strong indications given above for doubting that length in itself controls reading ease, all alterations of reading

passages prepared for this research preserve the original length of sentences as defined by the Dale-Chall formula. By so controlling length, the research study reduced the number of variables that had to be considered and eliminated the possibility that length, and not redundancy, might be considered by some readers to be the underlying variable in determining difficulty.

5. Review of the Literature: Cloze Procedure

What is required now is an operational definition of comprehensibility and a way of measuring it. Comprehensibility has already been discussed as the accessibility of the message in a text. Thus, it is a combination, in a way, of the ease or difficulty of the text itself and the reader's ability to get the message of the text. The Cloze procedure is an ideal instrument with which to measure comprehensibility because it has been shown to be sensitive to readability, the reader's ability to get the message, and also to redundancy.

The Cloze procedure was first defined and developed as a measure of readability by Wilson Taylor in 1953.[45] In the initial report of his research, Taylor introduced two important terms:

> A cloze unit may be defined as: any single occurrence of a successful attempt to reproduce accurately a part deleted from a 'message' (any language product) by deciding, from the context that remains, what the missing part should be.
>
> Cloze procedure may be defined as: a method of intercepting a message from a 'transmitter' (writer or speaker), mutilating its language patterns by deleting parts, and so administering it to 'receivers' (readers or listeners) that their attempts to make the patterns whole again potentially yield a considerable number of cloze units.[46]

The standard procedure developed by Taylor is outlined in his 1953 article. He prepared test passages by deleting every fifth word in a passage of 250 words and leaving blanks of a standard length. Subjects were asked to read the passage and to try to fill in the blanks. The passages were scored for the number of times the original words were replaced in the blanks. This total on each passage was considered a readability score, which could then be compared to scores on other passages treated in a similar way.

The other major component of Taylor's 1953 report is a comparison of Cloze as a measure of readability to the Flesch and Dale-Chall formulas. Several different kinds of experiments were carried out by Taylor to cross check the validity of his findings. He concluded from these experiments that the Cloze procedure was as good at ranking passages according to readability as the two formulas. Moreover, Cloze worked better than the formulas at predicting the readability of passages by Erskine Caldwell, Gertrude Stein and James Joyce. Finally, Taylor says:

> Potentially important, it seems, is the fact that a cloze score appears to be a measure of the aggregate influences of all factors which interact to affect the degree of correspondence between the language patterns of transmitter and receiver. As such, its potential usefulness is by no means confined either to readability or to the reading abilities of individuals.[47]

Thus, the usefulness of the Cloze procedure was recognized by Taylor and by others who read his report. A great deal more research followed this initial report.[48]

Ten years later, John Bormuth had carried out a series of experiments using Cloze procedure. Bormuth reported to the

International Reading Association in 1963 that Cloze was more accurate as a measure of readability and therefore superior to the formulas in measuring readability. He also found Cloze superior to the commonly used multiple choice test as a measure of reading comprehension. Bormuth sums up his findings as follows:

> 1. Cloze tests are valid and uniform measures of reading comprehension ability.
> 2. The cloze tests were valid and highly reliable predictors of the comprehension difficulties of the passages.
> 3. Cloze tests are appropriate for use with individuals and groups which vary widely in comprehension ability.[49]

By 1968, Bormuth had carried out still more research and reported the same conclusions with even greater confidence.[50] It is clear that the Cloze procedure has been very well validated as a measure of readability and reading comprehension.

The connections between the Cloze procedure and redundancy are discussed by Brendan Maher:

> The development of interest in the concepts of information and redundancy has led to an increasing use of the so-called Cloze technique for estimating redundancy in verbal utterances. It is a matter of common observation that normal utterances are redundant, that is to say, it is possible to eliminate parts of an utterance without impairing the ability of a listener to comprehend the message that the utterance was intended to convey. This redundancy rests, in turn, upon the fact that the probability that any given word will be followed by specific other words in normal speech is variable and for many words is higher than zero. Provided that we have the first few words of a sentence, then we can guess at the next word in the sequence with some real probability of being correct. The more probable it is that the next word will be a specific word, the more redundant its utterance at that point.[51]

The presence of the blanks in a Cloze passage provides a unique opportunity to study redundancy as the reader has used it. For each blank in the passage it seems possible to examine what readers put in the blank and to measure their use of redundancy in arriving at a particular fill-in item (or Cloze unit).

It is important to note that the Cloze procedure has already been shown to be sensitive to redundancy, particularly insofar as redundancy is related to language competence. The use and validity of Cloze for this purpose, as will be shown, is well documented. McLeod and Anderson used a Cloze test to measure the reader's contribution to the reduction of uncertainty in a passage,[52] and insofar as redundancy helps the reader to reduce his uncertainty about the message of the text, Cloze is shown in their study to be sensitive to redundancy.

Part of the reason that the Cloze procedure is an effective language testing device is explained by Bernard Spolsky.[53] He claims that part of one's knowledge of a language is the ability to understand messages with reduced redundancy such as Cloze passages and various other kinds of "noise" which decrease the redundancy of passages. Wendell Weaver and Albert Kingston report on the use of a factor analysis to study the relationship of Cloze tests to other language tests. Cloze tests related more closely to "redundancy utilization" or the recognition of redundancy in language than to verbal comprehension or to rote memory ability.[54]

In the context of using Cloze to measure language ability, the work of John Oller is most helpful. In several articles, Oller reports on research he has carried out using Cloze procedure with

adult students of ESL. He has carried out a number of experiments at UCLA which clearly show that Cloze tests are global or integrative tests. That is, Oller believes that Cloze tests measure overall language skills better than other proficiency tests and better than batteries of tests of specific language skills (aural comprehension, grammar, and the like).

Oller has also compared Cloze tests to discrete-point tests. Overall, he finds Cloze and other tests of integrative skills superior:

> In spite of the fact that some of the integrative skills tests seem to have little in common, and regardless of the fact that they may seem to be unreliable as far as scoring is concerned, repeated studies show that scores on tests of integrative skills tend to correlate better with teacher judgments, better among themselves, and better with other measures of language skills than do any of the discrete-point types because they more nearly reflect what people actually do when they use language.[55]

Thus Oller makes a strong case for the use of the Cloze procedure as a global measure of language skill. All of this documentation by Oller and the other scholars suggests that the Cloze procedure is a sound testing device for comprehensibility with the added attraction that it is sensitive to redundancy.

In using Cloze as a testing device, a critical problem to be resolved at the outset is how to arrange the blanks in test passages to answer the questions raised earlier about the nature of redundancy, its effect on comprehensibility, and the role of these matters in reading. McLeod and Anderson tried to deal with this problem by using a Cloze deletion formula of every eighth word.[56] This was not quite satisfactory, and they solved the problem by deleting one word

in six or eight and selecting for deletion words which were completely redundant for skilled readers. This work by McLeod and Anderson demonstrates that it is possible to manipulate the deletions in test passages without impairing the validity of the data. The methodology described in chapter 2 will make some use of McLeod and Anderson's solution.

The Cloze procedure has been widely used as a device to test various language features related to reading and the problems under discussion here. In particular, Cloze has been demonstrated to be a reliable measure of readability and comprehension, both features of the concept comprehensibility. In addition, Cloze provides a measure of the ease or difficulty of a text, and is sensitive to redundancy in a passage of text. These features of Cloze make it an ideal device for studying the relationship between redundancy and comprehensibility, a central focus of this project.

6. Summary

In the research study described in the following chapters, the literature on redundancy, readability and the Cloze procedure forms the base for the investigation. Scholars in psycholinguistics and reading theory have provided a clear description of the redundancy extant in a text. The research on readability describes ways in which we can predict how difficult a text will be for a particular reader, and in so doing indicates how a text must be controlled in order to study the effect of redundancy. The Cloze procedure provides an ideal test device for the problem at hand because of its ability to measure comprehensibility and its sensitivity to the reader's

linguistic competence and the redundancy in the text. The debt of this study to the research reviewed here should be abundantly clear. The study described carries all of this work one step further in an effort to discover whether or not the application of the operational definition of redundancy will serve to increase the comprehensibility of passages of text, as well as their inherent interest and writing quality.

FOOTNOTES: CHAPTER 1

[1]Frank Smith, Understanding Reading (New York: Holt, Rinehart and Winston, 1971), p. 9. Hereafter cited as UR, with a page number referring to this edition.

[2]UR, p. 19.

[3]UR, p. 20.

[4]UR, p. 20.

[5]UR, p. 220.

[6]UR, p. 135.

[7]Kenneth Goodman, "Reading: A Psycholinguistic Guessing Game," in Current Topics in Language: Introductory Readings, ed. Nancy Ainsworth Johnson (Cambridge, Massachusetts: Winthrop Publishers, 1976), pp. 370-383.

[8]E. Brooks Smith, Kenneth S. Goodman and Robert Meredith, Language and Thinking in the Elementary School (New York: Holt, Rinehart and Winston, 1970), p. 251.

[9]Smith, Goodman and Meredith, pp. 257-58.

[10]Goodman, pp. 375-76.

[11]Colin Cherry, On Human Communication, 2nd ed. (Cambridge, Massachusetts: MIT Press, 1966), p. 182.

[12]Cherry, pp. 120-21.

[13]Cherry, p. 19.

[14]John R. Pierce, Symbols, Signals and Noise (New York: Harper and Brothers, 1961), p. 143.

[15]Wendell Garner, Uncertainty and Structure as Psychological Concepts (New York: John Wiley and Sons, 1962), p. 215.

[16]UR, p. 21.

[17]UR, p. 22 is a detailed discussion of distributional redundancy.

[18]UR, p. 134.

[19]See Garner, p. 220f. for a discussion of Shannon's guessing game technique, multivariate analysis, and so on. At best, these statistical analyses provide only lower-bound estimates of redundancy and do not solve the problems under discussion here.

[20]Garner, p. 239.

[21]Garner, p. 220.

[22]Garner, p. 220.

[23]Garner, p. 220.

[24]Garner, pp. 220-21.

[25]Garner, p. 224f.

[26]Garner, p. 239.

[27]Garner, p. 221.

[28]Frank Smith, personal communication, March, 1976.

[29]Garner, p. 281.

[30]Garner, p. 242.

[31]Robert Ruddell, "Reading Comprehension and Structural Redundancy in Written Material," in Reading and Inquiry, ed. J. A. Figurel (Newark, Delaware: International Reading Association, 1965), pp. 308-11.

[32]Ruddell, "Reading Comprehension. . .," p. 309.

[33]Edgar Dale and Jeanne Chall, "The Concept of Readability," Elementary English, 26 (January 1949), p. 23, cited in Jeanne Chall, Readability: An Appraisal of Research and Application (Columbus, Ohio: Bureau of Educational Research, 1958), p. 7.

[34]John Bormuth, "Introduction," in Readability in 1968, ed. John Bormuth (Champaign, Illinois: NCTE, 1968), p. v.

[35]Bormuth, pp. 3-4.

[36]George Klare, "Assessing Readability," Reading Research Quarterly, 10, No. 1 (1974-75), pp. 62-102.

[37]Rudolf Flesch, "A New Readability Yardstick," Journal of Applied Psychology, 32 (June 1948), pp. 221-33. The original formula was developed in 1943, as noted on p. 221 of the article.

[38]Edgar Dale and Jeanne Chall, "A Formula for Predicting Readability," Educational Research Bulletin, 27 (January 21 and February 18, 1948), pp. 11-20, 28, 37-54.

[39]Flesch (1943) defines these as "All nouns with natural gender; all pronouns except neuter pronouns; and the words people (used with a plural verb) and folks." p. 223.

[40]Flesch (1943) defines these as spoken sentences, direct or indirect, sentences addressed to the reader, exclamations and incomplete sentences whose meaning must be inferred. p. 223.

[41]Klare, p. 70.

[42]Dale and Chall, "Formula," p. 45f.

[43]Klare, pp. 62-102.

[44]Klare, p. 97.

[45]Wilson L. Taylor, " 'Cloze Procedure': A New Tool for Measuring Readability," Journalism Quarterly, 30 (1953), pp. 414-33.

[46]Taylor, p. 416.

[47]Taylor, p. 432.

[48]See, for example, Wilson L. Taylor, "Recent Developments in the Use of 'Cloze Procedure,' " Journalism Quarterly, 33 (1956), pp. 42-48 and " 'Cloze' Readability Scores as Indices of Individual Differences in Comprehension and Aptitude," Journal of Applied Psychology, 41, No. 1 (1957), pp. 19-26.

[49]John Bormuth, "Cloze as a Measure of Readability," in Reading as an Intellectual Activity, ed. J. Allen Figurel (New York: Scholastic Magazines, 1963), p. 134. (International Reading Association Conference Proceedings, Volume 8.)

[50]Bormuth, Readability, p. v.

[51]Brendan Maher, "Language and Psychopathology," in Communication, Language and Meaning, ed. George Miller (New York: Basic Books, Inc., 1973), p. 257.

[52]John McLeod and Jonathan Anderson, "An Approach to Assessment of Reading Ability Through Information Transmission," Journal of Reading Behavior, 2, No. 2 (Spring, 1970), pp. 116-43.

[53]Bernard Spolsky, "Reduced Redundancy as a Language Testing Tool," ERIC, ED 031702 (September 1969), 17pp.

[54]Wendell Weaver and Albert Kingston, "A Factor Analysis of the Cloze Procedure and Other Measures of Reading and Language Ability," Journal of Communication, 13 (1962), pp. 252-61.

[55]John W. Oller, Jr., "Discrete Point Tests versus Tests of Integrative Skills," in Focus on the Learner, ed. John W. Oller, Jr. and Jack Richards (Rowley, Massachusetts: Newbury House, 1973), p. 198.

[56]McLeod and Anderson, p. 119f.

CHAPTER 2

METHODOLOGY

1. Introduction

The methodology for the formal research study was devised after two pilot studies were carried out. In general, much of the methodology is the same for both the pilots and the formal study. For this reason, the pilot studies are only briefly sketched below, and are discussed thereafter only where they provide the rationale for, or some further insight into, the methodology. After a brief description of the pilots, this chapter describes the formal study in several sections, each dealing with one aspect of the study: the materials, hypotheses, subjects and so on, with reference to the pilot studies as necessary.

2. The Pilot Studies

The purpose of the pilots was to test the materials and procedures to be used in the formal study. The first pilot revealed several minor problems, most of which were corrected before the second pilot was run. While the pilots did not exactly replicate the main study, they were sufficiently close to the main study to reveal problems in design before the main study was conducted.

The subjects in the pilot studies were students in English 0150 (Freshman Composition) at Wayne State University in the fall

quarter of 1976 and the early part of the winter quarter of 1977. A total of 157 readers were involved in the first pilot, while 123 were involved in the second pilot.

A total of twelve passages on three different topics was used in the pilot. They were prepared as Cloze tests[1] using a one-in-six deletion formula and choosing for deletion only nouns and main verbs. The three topics were: first, the origin and development of male dominance in society (100 series),[2] second, the effect of radio and television on sports (200 series),[3] and finally, the evaluation of the performance of government (300 series).[4] Reasons for these choices are described below. Each original passage was rewritten in three variant forms: a form with syntactic redundancy increased, a form with semantic redundancy increased, and a form with both syntactic and semantic redundancy increased. Parameters for rewriting are also described below. The major controls on the materials were total length for each passage (about three hundred words), readability level as measured by the Dale-Chall formula,[5] the total number of blanks (about fifty), and the numbers of nouns and main verbs deleted. The words deleted were nearly the same in all forms of each passage.

Subjects for the pilot were solicited by asking members of the English Department teaching English 0150 to volunteer their classes. Each instructor who volunteered was given a packet of test passages and a set of instructions for administering the test. Students read the passages during the ninth and tenth weeks of the fall term and in the early part of the winter term. Subjects were asked to fill in an information form, the main purpose of which was to find out their class level and whether they were native speakers

of English. Other background information was also requested, such as age, sex, major, and so on. In the pilots, non-native speakers were eliminated from the eligible subject pool. The subjects were given twenty minutes to read the passage, fill in as many blanks as they could and to answer the two rating questions, concerning interest and writing quality of the passage. The forms were then collected by the instructor and returned for scoring. This administration procedure was revised slightly in the formal study.

The passages were scored by an acceptable word scoring system such that the exact word as the original, any close form of the word (a third person singular present tense verb with the -s marker omitted, for example), or any close synonym of the word was counted as a correct response. This was a simple right/wrong scoring system. In the formal study, a more detailed scoring system was adopted and is described below.

The pilots were helpful in establishing a workable methodology for the formal study. Minor problems with the deletion pattern in one of the series, with the administration procedure and with the content and form of the background and rating questions were identified and solved before the formal study was conducted. Other insights from the pilots are discussed below.

3. Formal Study

3.1 Materials

The materials used were the same for the formal study as those used in both pilots. The materials were chosen by following several guidelines, having to do with topics, difficulty level and subject

familiarity and interest.

The first guideline dealt with topics. Material was chosen from three disparate areas in order to ensure that there would be an equal possibility that a chosen topic might fall within a subject's area of interest. For this reason, the general areas of psychology, sports and political science were selected. The issues discussed in each passage are of a rather general nature.

The second area of concern in choosing material was the difficulty level of the material. Since the genesis of this project had to do with the question of how to make difficult material easy to read, relatively difficult material was chosen. In the materials chosen, readability is measured and controlled by the classic Dale-Chall readability formula, where readability is determined principally by sentence length and vocabulary level. All the materials used are scored by the formula as appropriate for high school seniors or college students. Within this level of difficulty, material was chosen where the author's main ideas seemed clear, and where at least one complete idea was discussed in the three hundred word text chosen.

A third area of concern in choosing the material was that it be material the subjects could be expected not to have seen before. At the same time, however, there was an attempt to choose topics which any adult might be interested in and/or concerned about. These factors controlled the choice of sources and topic areas. In the pilots, some subjects indicated an interest in each of the topics, which suggested that the guidelines for selecting the material worked reasonably well.

3.2 Guidelines for Preparation of Passages

Each passage was prepared as a Cloze test in four forms or versions. The first version (01) is the original text. The second version (02) is the form which incorporates increased syntactic redundancy. The third version (03) is the form which incorporates increased semantic redundancy. The fourth version (04) is a combination of the second and third versions, with increased syntactic and semantic redundancy. In general, the rewritten versions were meant to look and sound like unadulterated prose (so that, for instance, the use of the repetitive phrases with or from a preliminary study was abandoned), and the traditional factors in readability--sentence length and vocabulary level--were kept relatively constant. On all versions of a passage, the readability level (Dale-Chall score) did not vary more than .5 from the original. Where this score did vary at all, it went up (i.e. difficulty increased) but not more than .5 in any case.

There were three specific groups of rewriting guidelines that were followed for the versions with increased redundancy. The first group of guidelines was a set of general rules followed closely on all forms: 1) all rewritten forms may not vary more than five words from the original in total length; 2) the last sentence may be edited for length considerations, but the content must remain intact; 3) the readability score may not vary more than .5 from the original; 4) the semantic changes should be spread through the passage; 5) changes which would produce awkward or nonsensical sentences are not made.

A second group of guidelines dealt with strategies for increasing syntactic redundancy. These are: 1) rearrange highest sentence (in a sentence with embeddings) to NP VP order;[9] 2) replace recoverable subjects and verbs; 3) replace pronouns with their referents; 4) add a referent to any unclear "this"; 5) replace dummy subjects such as "it" and "there" with real subjects in the highest sentence or any embedded sentence; 6) change passives to actives. These changes were applied in every case where the criteria for a change existed, except where awkward or nonsense phrasings resulted. The number of changes made was not numerically controlled.

The third group of rewriting guidelines was designed to increase semantic redundancy. Since these were additions to the text, a numerical limit of no less than three, but no more than five, semantic changes was imposed. The limit was imposed principally because it would have been impossible to control for length otherwise. Again, the general guideline that the text must not sound unnatural was important in selecting the place and nature of the semantic revisions. In general, these additions serve one of the following purposes, and each takes on its respective form: 1) the addition of an example, marked by a phrase beginning "for example" or "for instance"; 2) the specification of particulars, marked by a phrase beginning with "such as"; 3) the clarification of a term or concept with a defining phrase set off by parentheses but no verbal marker. From these choices, in general, three different additions were made on the version with increased semantic redundancy, using, again, the general guideline of "normal prose."

3.3 Deletions

With the parameters for the choice of the topics of the passages and the guidelines for rewriting, the passages were controlled, as noted, for length and readability level. Also controlled were numbers of nouns and main verbs deleted (nouns were preferred; verbs did not exceed twenty percent of deletions), and the number of blanks per passage (about fifty, using a one-in-six deletion formula). Finally, despite the rearrangement and addition in the 02 (with increased syntactic redundancy) and 03 (with increased semantic redundancy) forms, and the combination of these factors in the 04 (with both syntactic and semantic redundancy increased) forms, an effort was made to delete the same words in all versions. Approximately 80% of the items deleted in each version were identical.

3.4 Hypotheses

One major hypothesis and two corollary hypotheses were investigated in this study. The major hypothesis is that increased redundancy yields increased comprehensibility. The two corollary hypotheses are, first, that increased redundancy yields increased interest in a text, and second, that increased redundancy improves the subjective evaluation of the quality of writing in a text.

The major hypothesis is suggested by much of the literature. Smith and Goodman both indicate, for example, that even in unmodified text, readers rely on the inherent redundancy of the language to help them get meaning at maximum speed. Additional redundancy, inserted artificially, seems likely to increase the availability of the meaning especially in difficult material.

The corollary hypotheses are directly related to the major hypothesis. If the meaning becomes more accessible to readers by virtue of increasing redundancy, their interest in the text is likely to increase as they get more meaning more easily. The presence of additional redundancy, then, should increase the reader's interest in the text. The first of the two corollary hypotheses investigates the relationship between increased redundancy and reader interest.

The second corollary hypothesis calls for a subjective evaluation on the part of the subject. As with the first corollary, if meaning becomes more accessible through added redundancy, the subject may perceive the passage as being well-written. The criteria given to the subject are clarity, organization and style. Each of these items may seem to improve as meaning becomes more available to the reader. The second corollary hypothesis, that increased redundancy improves the subjective evaluation of the quality of writing in a text, is designed to study the relationship between increased redundancy and writing quality as perceived by the subject.

Redundancy was defined and measured in terms of the series of guidelines used in rewriting the materials. Comprehensibility was measured by the Cloze test scores, using an acceptable word scoring system, which is described below in the section on scoring. Interest was measured with two questions. The first was a pre-reading measure, answered by subjects as part of the background information form, which asked the subjects to rank the following five topics in order of their interest in reading about them: political science, famous people, sports, women's liberation and science. This question is referred to below as the "ranking question." Interest was also

measured by a post-reading interest rating question, which asked: "Despite your ranking of the topics above, how interesting did you find this passage?" Five answers were provided to the subject, who was asked to circle one choice: a) very interesting, b) moderately interesting, c) neither interesting nor boring, d) moderately boring, e) very boring. Writing quality was measured by a second rating question. In both pilots, subjects were asked how they would grade the passage on an A to E scale, if they were English teachers. Some subjects clearly did not have the criteria to make such a judgment, so the following question was substituted as a measure of writing quality: "In terms of clarity, organization, and style, how well written do you think this passage is?" Subjects were asked to circle one choice from the following: a) very well written, b) moderately well written, c) neither good nor bad, d) moderately poorly written, e) very poorly written.

3.5 Subjects

The subjects in the formal study were 240 freshman native speakers of English enrolled in the English 0150 (Freshman Composition) classes during the fifth and sixth weeks of winter term, 1977, at Wayne State University. The subjects were solicited by using the same volunteer procedure as was used to obtain subjects for the pilots. Since upperclassmen and non-native speakers of English also enroll in English 0150, the background information form was used to eliminate ineligible subjects. Since all the versions of the three passages were randomized, data collection continued until twenty eligible subjects had read each version. Each subject read only

one passage.

This methodology for obtaining eligible subjects was an improvement over the pilot studies. In the pilots, non-native speakers of English were eliminated from the eligible subject pool, but upperclassmen remained, and no effort was made to equalize the number of subjects reading each passage in the pilots. The reason for this was that the pilots were only an attempt to make sure the study would work smoothly and provide suitable data. The results of the pilots were not subject to detailed statistical analysis, and even if they had been, would not have provided sound data due to a failure to randomize the forms and other minor problems.

3.6 Procedures

In the formal study, unlike either pilot, all versions of all passages were randomized using the standard procedure[7] and the random numbers table.[8] In this way, each subject received one of the twelve versions at random, controlling a large variety of environmental factors such as time of day, teacher influence, program background (i.e., a number of subjects with similar course schedules from a particular program such as pharmacy tend to take a particular class), classroom noise, light and heat levels, and so on. Data collection continued until twenty eligible subjects had read each of the twelve passages.

Some parts of the administration procedure followed in the pilot studies did not require much revision. The directions given to the instructor and to the subjects were apparently sufficiently clear. The instructor's directions dealt with completing the

background information form and timing the exercise. The instructor was also authorized to explain this project in a very general way and to tell the subjects that the exercise did not affect their grades or academic careers. The instructor did not know which forms were rewritten, and which were not, and neither did the subjects. Subject directions were given just above the passage itself, and allowed the subject to reread the passage as much as necessary to fill in the blanks. The subject was asked to try to fill in all the blanks, and to put only one word in each blank.

The background information form was designed to eliminate ineligible subjects, and to collect other data which was considered potentially relevant to the subject's performance on the Cloze test.[9] The key eligibility questions were the class level of the subject and the language spoken at home. The other data collected included the following items: the subject's sex, age, major or area of academic interest, grade point average, degree sought, parents' average yearly income, favorite leisure time activity, and the languages studied besides English.

The subject was also asked to answer two attitudinal questions. The first of these was a general reading attitude question, which asked how often the subject read for pleasure, and gave four choices ranging from "every day" to "rarely." The second attitude question was the ranking question which asked the subject to rank five topics, (including the three which were the topics of the passages in the study), in terms of which he would most like to read about. All these questions were answered prior to the twenty-minute timed Cloze exercise.

Twenty minutes seemed to be an appropriate and sufficient amount of time for most subjects to read the passage and fill in the blanks. In the first pilot, subjects were also asked to answer the post-reading rating questions within the time limit. As a result, many subjects did not answer the rating questions. Therefore, the twenty minute limit on the reading was maintained, but subjects were not asked to answer the rating questions until the timed session ended. This produced more complete answers to the rating questions on the second pilot and in the formal study. The revision, as previously discussed, of the second rating question dealing with writing quality, also improved the ratings data.

Briefly, in the formal study the subjects received the passages and directions from their instructor. Each subject first answered a series of background questions, dealing with class level, major, native language, citizenship, age, sex, and so on. Each subject then had twenty minutes to read the passage and fill in the blanks. Assignment of passages to subjects was random. The subject was asked to put only one word in each blank and to attempt to fill every blank. Subjects were permitted to reread as often as necessary during the twenty minutes. At the end of the permitted time, all subjects were asked to answer the two rating questions. The passages were then collected and returned for scoring.

3.7 Scoring

The data scoring followed a set of guidelines developed in part in response to the pilot data. In general, this is an acceptable word scoring system,[10] such that 1-4 are considered right and 5-7

and 0 are considered wrong. The validity of this type of scoring is well-established. The system used was made more detailed to make possible more complex analyses if needed. The scoring system used to score the data is presented in Table 1. The material from the information sheet, including the post-reading ratings of interest and writing quality were coded by a numerical system to make the data readily accessible in a computer file. The scored data were checked for reliability by a second scorer[11] and filed in the Wayne State University computer for analysis.

3.8 Analysis

The data were analyzed using three statistical procedures: the analysis of variance on the Cloze test results and regression and the correlation ratio to study the relationship of the Cloze scores to the other variables examined in the study. The purpose of the analysis was principally to see whether adding redundancy, either syntactic or semantic or a combination of both types, would significantly improve the Cloze test scores. The analysis of variance is the procedure of choice for this type of analysis, since four mean scores (one for each version of a passage) are compared in analysis of variance to discover significant differences. The further purpose of the analysis was to see whether adding redundancy created increased interest and improved judgments of writing quality. For these issues, regression was chosen since it can reveal a relationship of each factor to the Cloze scores. Finally, the analysis included an attempt to account for an anomaly in the Cloze score results and to examine the role of other factors in the study. These

TABLE 1

DATA SCORING SYSTEM FOR THE MAIN STUDY

0 = all blanks.

1 = exact match. Allow also minor variations in spelling such as dominence for dominance.

2 = high acceptable. Allow such things as verb forms without a tense marker, number marker, etc. Also count as 2 alternate noun forms, like dominance for domination. In general, these should be near to exact matches, or alternate forms of tbe original word. To count as 2, the item must be the same part of speech as the exact answer.

3 = mid acceptable. Allow any close synonym which is the same part of speech as the original word. Use context to judge only part of speech, not syntactic acceptability. Use synonymy with exact item to determine right or wrong.

4 = low acceptable. Allow any close synonym which is a different part of speech than the original item. 4s should be close synonyms, but not syntactically the same as the exact item. Use 4 also for cases where the filled in answer is a synonym but part of speech is unclear.

5 = low unacceptable. An answer which is not synonymous, but which is the same part of speech as the original item. For a 5, ignore the presence or absence of plural, and tense markers.

6 = high unacceptable. An answer which is not synonymous, and which is not the same part of speech as the original item.

7 = wrong. An answer which is not synonymous and for which part of speech is unclear.

Note on pronouns: if antecedent was 1, 2, 3 or 4, score a correct pronoun as 3. If antecedent was incorrect, score the pronoun as a 6.

other factors were considered by a combination of further regression and the use of the correlation ratio. The results of these analytic procedures are reported in detail in chapter 3.

4. Summary

This chapter has discussed the methodology used in the formal study. The methodology included the preparation of the reading materials, the procedure for administering the materials, a description of the subjects who read the materials, the scoring method used on the Cloze tests and the analyses carried out on the data. The two pilot studies carried out prior to the formal study have been briefly sketched here and discussed as they influenced the design of the formal study. The results of the pilot studies and the materials used in both the pilots and the formal study appear in the appendix.

FOOTNOTES: CHAPTER 2

[1]Wilson L. Taylor, " 'Cloze Procedure': A New Tool for Measuring Readability," Journalism Quarterly, 30 (1953), pp. 414-33.

[2]Alfred Adler, Understanding Human Nature (New York: Fawcett World Library, 1954), p. 105. Adapted.

[3]Leonard Koppett, New York Times Guide to Spectator Sports (New York: New York Times Publishing, 1976), pp. 1-2. Adapted.

[4]Roy C. Macridis and Robert E. Ward, Modern Political Systems: Europe, 2nd ed. (Englewood Cliffs, New Jersey: Prentice-Hall, 1968), pp. 25-26.

[5]Edgar Dale and Jeanne Chall, "A Formula for Predicting Readability," Educational Research Bulletin, 27 (January 21 and February 18, 1948), pp. 11-20, 28, 37-54.

[6]Precedent for this was set by Robert Ruddell, "Reading Comprehension and Structural Redundancy in Written Material," in Reading and Inquiry, ed. J. A. Figurel (Newark, Delaware: International Reading Association, 1965), pp. 308-11. (This text is also IRA Proceedings, Vol. 10.)

[7]Henry Klugh, Statistics: The Essentials for Research (New York: John Wiley and Sons, 1970), p. 86.

[8]Klugh, pp. 338-39.

[9]Roger Farr, Reading: What Can Be Measured (Newark, Delaware: International Reading Association, 1969), pp. 1-32.

[10]Robert Ruddell, "A Study of the Cloze Comprehension Technique in Relation to Structurally Controlled Reading Material," in Improvement of Reading Through Classroom Practice, ed. J. A. Figurel (Newark, Delaware: International Reading Association, 1964), pp. 298-303. (This text is also IRA Proceedings, Vol. 9.)

[11]Ms. Barbara S. Reetz served as second scorer.

CHAPTER 3

THE RESULTS

1. Introduction

The results reported here are derived from the main research study which was carried out during the winter quarter of 1977 at Wayne State University, using the methodology described in chapter 2. The major results reported are the results of the analysis of the Cloze test scores. These scores were checked for reliability and subjected to an analysis of variance. Two other types of analysis were also carried out: regression and the correlation ratio.

The analysis was carried out for the obvious purpose of providing evidence for conclusions about the hypotheses under study. The central question to be answered is, does adding redundancy, as it has been defined, make a difference in comprehensibility, interest and/or writing quality? Data analysis was also undertaken to study the relationship between the variables studied and the Cloze score patterns. At the outset, it was not clear which variables might prove to be significant, and the detailed analysis of all variables was carried out to determine which factors, if any, had affected the Cloze scores.

In addition, however, because some of the results are anomalous, the data analysis represents an investigation of several explanations for the anomalous results. Attention must be focused not only on

the hypotheses, but also on whether any of the other variables as they were investigated can shed light on the results obtained. The discussion of the results is intended to provide evidence for the conclusions reached in chapter 4, and also to investigate possible explanations for the anomalous data which appeared in the study.

2. Cloze Test Results

The Cloze test results are divided into two parts: the reliability of the scoring and the analysis of variance. The reliability of the scoring of the principal investigator was checked by the use of a second scorer, who was informed about the nature of the study in a very general way. She was given a copy of the scoring system and several passages from the pilot were scored simultaneously by both scorers until the second scorer felt confident of her judgment. The second scorer was then given a total of thirty-six passages to score, independently. Passages completed by six subjects chosen at random were taken from each of the following forms: 101, 104, 201, 204, 301 and 304. The passages were chosen at random from the eligible passages, and the form number on each test sheet was concealed from the second scorer. The second scorer did not know which forms were altered and which were not, nor the specific nature of the text modifications.

Table 2 presents the results of the reliability check, which was done by comparing the scores of the principal investigator to those of the second scorer using the Pearson correlation coefficient. The Pearson test is a statistical procedure for determining both the direction and the degree of agreement between two groups of numbers--

here, two groups of scores. The average number of differences for each form, the number of blanks and the resulting percentage of agreement are also presented.

TABLE 2

PEARSON CORRELATION OF RELIABILITY

Form	Avg. # of Diffs./ Number of Blanks	% Agree	Correlation Coefficient
101	4.67/48	90.27	.91
104	4.5 /48	90.62	.89
201	4.33/50	91.30	.97
204	2.33/50	95.34	.93
301	2.67/50	94.66	.98
304	2.83/49	94.22	.95
ALL			.96

To do a complete correlation study, the scores were compared first on each individual form. So, on the first form, 101 (the original text on male dominance in society), the correlation of the two sets of six scores is .91. Because the sign of the correlation is positive, it indicates that the two scorers agreed. Because the correlation is relatively close to 1.00, it is considered a high correlation. The other five forms were handled in the same way, with the correlations as indicated in the last column of Table 2.

The bottom line of the table indicates the overall correlation on all thirty-six forms involved. This figure is not an average of the correlation on each form, but was obtained by correlating the two sets of scores on all thirty-six forms treated as a single group. An overall correlation of .96 indicates high agreement between the

principal investigator and the second scorer. This correlation was done before going on to the scoring of all forms, and although a second scorer was not used at the end of the process, the principal investigator's judgment is established as reasonably objective and we can assume accuracy on all scoring.

The second part of the Cloze test results is the analysis of variance which was carried out on the Cloze scores of all 240 subjects in the formal study. The analysis of variance is the appropriate statistical procedure to use when more than two means are being compared. In this case, the mean scores of the 02, 03 and 04 forms were being compared to the 01 in each of the three series. For the purpose of analysis, fill-ins coded or scored as 1, 2, 3, or 4 were considered right and those scored 5, 6, 7 or 0 were considered wrong.

The analysis of variance focuses first on the two major effects in this study: syntactic and semantic. The procedure involves comparing the scores obtained in the presence of the variable or effect to those obtained in the absence of the variable. Thus to examine the effect of syntactic modification, in the 100 series, the analysis of variance compares the means of forms 101 and 103 (with no syntactic modification) to the means of forms 102 and 104 (with syntactic modification). An identical procedure is used for the other series and for the semantic variable as well. Thus, the analysis of variance does not indicate exactly whether the difference between the mean score on form 101 is significantly different from the mean on form 102, but does show whether the modification in question makes a significant difference in the scores.

Table 3 shows the mean, the standard deviation and the mean percentage scores of the Cloze tests by form, as well as the results of the analysis of variance. The analysis of variance also considers the two-way interaction of the two main effects examined in the study. The result is reported as the "2 way" effect in the table. The two-way interaction is in essence an analysis of the 04 score in each series as compared to the 01. Here, the results are quite interesting. When both types of modification are used, the Cloze score is significantly higher (where significance is greater than the level of .01) in two of the three series. In the third case, both types of modification make no significant difference in the Cloze score, that is, the differences observed in the means on the 200 series are not significant.

Because of the way the analysis of variance works, it is important not to be distracted by the mean scores for each form. The means indicate only the direction of significant differences created by the presence or absence of an effect: an effect may raise means or lower them, but changes may or may not be significant.

The significance of the differences is really the crucial point. Significance is being used here in its technical sense, that is, to indicate that a significant difference is one which we can statistically predict will hold for the whole <u>population</u> from which the sample was drawn. The greater the significance of an F-ratio in an analysis of variance, the more accurately we can predict how the whole population would score.

Turning back to Table 3, we can see that in the 100 series (on male dominance), the addition of syntactic redundancy alone

TABLE 3

CLOZE TEST MEANS, STANDARD DEVIATION AND
THE ANALYSIS OF VARIANCE

Cloze Test Results

Form	Mean	Std. Dev.	Pct.
101	34.45	5.943	71.77
102	27.80	6.764	56.73
103	30.85	5.163	69.96
104	39.00	7.167	81.25
201	24.90	8.296	49.80
202	21.75	6.843	43.50
203	20.40	6.832	40.80
204	20.05	4.639	40.10
301	15.75	7.454	31.50
302	17.00	7.921	34.00
303	13.45	6.194	26.90
304	24.90	6.836	49.80

Analysis of Variance

Series	Effect	F	Sig. of F
100	Syntactic	.283	.999
	Semantic	7.261	.008
	2 Way	27.537	.001
200	Syntactic	1.333	.251
	Semantic	4.182	.042
	2 Way	.853	.999
300	Syntactic	15.860	.001
	Semantic	3.084	.079
	2 Way	10.230	.002

raises the means (compare 101 and 103 to 102 and 104) but this increase is not significant (where significance is greater than .01). For the 200 series (on sports) the mean goes down, but not significantly. Only in the 300 series (on government performance) does syntactic redundancy significantly improve the Cloze scores. The addition of semantic redundancy alone makes no significant difference in the 200 series although the means drop. In the 300 series, there is a non-significant increase in the mean, while in the 100 series the addition of semantic redundancy significantly improves the mean Cloze score.

The two-way interaction in the analysis of variance shows a different pattern than that noted for either effect alone. The two-way interaction compared the exact combined presence of both effects (the 04 in all series) to the absence of the effects (the 01 or original in all series). Here, there is a significant improvement in the means in the 100 and 300 series, and, although the mean in the 204 drops as compared to the 201, the change in the 200 series is not significant.

It is important to note that the 200 series is anomalous in all cases. Neither type of redundancy alone nor the combination of both types of redundancy makes any significant difference to the means in the 200 series.

3. Anomalous 200 Series Analysis

In view of these results, an obvious question arises. Why doesn't the double modification make a significant difference in the 200 series? This sort of question about the 200 series has plagued

this project from the beginning, and is ultimately left unresolved.

Two types of analysis were carried out in order to explain the results of the 200 series: regression and correlation ratio. The outcome of this further analysis is only suggestive of possibilities. No clear explanation for the 200 series can be given at this time. It may seem odd that both regression and the correlation ratio were used in the further analysis of the data. The rationale for using both procedures--one on some variables, and the other on others--becomes clear when the nature of the variables studied and the data they produced is made clear. The variables investigated were chosen for one of two reasons: either they were expected to provide evidence for a corollary hypothesis or they were known or suspected to be factors which might influence a subject's performance on the Cloze reading task. Variables chosen for the first reason included a pre-reading measure of interest on five topics, including the three involved in the study, a post-reading measure of interest in the passage read, and a post-reading evaluation of the writing quality of the passage read. Variables chosen because they might influence reader performance included: age, sex, grade point average, annual family income, major, degree sought, leisure time activity, number and type of languages studied and a reading attitude question.

This list of variables falls into three major categories in terms of the scale on which each item can be measured. Some of the variables produce ordinal scale data. This type of data is that which is in a numerically ordered form. Raw test scores like the Cloze scores are ordinal, as are the data on age, grade point average and income. Many of the other items are nominal scale data, i.e.

that which is arranged in discrete categories which have no numerical order or progression. The data on sex, for example, are nominal for this reason, as are the data on major, degree, leisure time activity, the number and type of language studied, and the pre-reading ranking question. A third category of data is that which is really nominal but which can be treated as ordinal because the categories represent a progression in value (this is like the data produced by Likert-type scales).[1] The data in this category include the post-reading measures of interest and quality, and the reading attitude question.

Thus, to investigate the 200 series Cloze scores further, and also to obtain results on the two corollary hypotheses, a correlation of the ordinal Cloze scores and other ordinal, nominal and mixed data was in order. The best analysis for this type of data, when the relationship can be expected to be linear, is regression.

Regression is a procedure which predicts the score on one variable from the scores on one or more other variables in the study. Such an analysis would show the relationship (if any) between the Cloze scores and the other variables. Also, regression has as its special virtue the capacity to look at <u>all</u> of the variables in the study and to rank them in order of how well they predict the Cloze score. The key problem with the use of regression is that it provides a meaningful analysis only when all the data are in a numerically ordered form (ordinal). Regression could thus only be used with the following variables: age, income, grade point average, reading attitude, interest and quality.

The income variable was excluded from the regression analysis because of a specific problem with regression and the income data.

In the SPSS computer program for regression,[2] the computer must reject from the entire procedure any case (subject) which contains missing data on any of the variables subject to the regression analysis. Many subjects did not respond to the income question, and in order to avoid having the regression done on a relatively small number of complete cases, income was omitted. All methods of restoring the income data or obtaining it from other sources at Wayne State University were exhausted without success.[3]

The regression analysis was carried out on the variables of age, grade point average (GPA), reading attitude (READ), interest and quality (QUAL). This procedure attempts to predict the Cloze score from each variable involved. The variables are selected in a step-wise order, such that for each form, the variable which best predicts Cloze score is chosen first in the regression. Each of the remaining four variables is added in, one step at a time, and with each addition, a new multiple R correlation is calculated to show the resulting improvement in the prediction. Variables are calculated in this manner until a tolerance level is reached, at which point there is no improvement in the prediction, and then no further calculation is done. In some cases, this tolerance level was reached after the third variable, so no further calculation is done. Also, for each variable, a simple R is calculated for that variable alone, and the significance of the multiple R is provided. The results of the regression analysis appear in Tables 4, 5, and 6.

The results of the regression on the post-reading ratings of interest and quality will be ignored temporarily. They are discussed in detail below. The explanation sought here is one that will account

TABLE 4

RESULTS OF REGRESSION ANALYSIS ON
SELECTED VARIABLES-100 SERIES

Variable	Form	Step Order	Simple R	Multiple R	Sig. of Multiple R
Interest	101	3	.08589	.60643	NS
	102	1	.47511	.47511	.05
	103	2	.12623	.24072	NS
	104	2	-.35409	.48519	NS
Quality	101	1	-.43877	.43877	NS
	102	3	.35871	.66206	.05
	103	4	.02471	.29846	NS
	104	1	.41712	.41712	NS
GPA	101	2	.37090	.60050	.05
	102	2	.47272	.65111	.01
	103	1	.19802	.18902	NS
	104	4	.05028	.53906	NS
READ	101	4	.11504	.60939	NS
	102	4	-.23326	.66242	NS
	103	3	.18066	.27883	NS
	104	5	-.26633	.54904	NS
AGE	101	not calculated			
	102	not calculated			
	103	not calculated			
	104	3	.24107	.51271	NS

Cloze %	
101	71.77
102	56.73
103	69.96
104	81.25

TABLE 5

RESULTS OF REGRESSION ANALYSIS ON
SELECTED VARIABLES-200 SERIES

Variable	Form	Step Order	Simple R	Multiple R	Sig. of Multiple R
Interest	201	4	-.42857	.65564	NS
	202	2	-.23906	.42549	NS
	203	2	-.12912	.59350	.05
	204	1	-.24082	.24082	NS
Quality	201	1	-.50233	.50233	.05
	202	3	.08645	.47938	NS
	203	4	.21291	.63950	NS
	204	2	.16107	.39103	NS
GPA	201	5	.43699	.65897	NS
	202	not calculated			
	203	1	.42899	.42899	NS
	204	3	.19883	.46240	NS
READ	201	2	-.33342	.55713	.05
	202	not calculated			
	203	5	-.31085	.65309	NS
	204	5	.11348	.48786	NS
AGE	201	3	.15664	.62063	.05
	202	1	-.36754	.36754	NS
	203	3	-.08635	.61378	NS
	204	4	.20288	.48464	NS

Cloze %	
201	49.8
202	43.5
203	40.8
204	40.1

TABLE 6

RESULTS OF REGRESSION ANALYSIS ON
SELECTED VARIABLES-300 SERIES

Variable	Form	Step Order	Simple R	Multiple R	Sig. of Multiple R
Interest	301	1	-.40765	.40765	NS
	302	3	.19617	.73828	.01
	303	4	-.02863	.49953	NS
	304	3	.25078	.58333	NS
Quality	301	4	-.16118	.52591	NS
	302	2	-.48140	.70147	.01
	303	1	-.22300	.23300	NS
	304	not calculated			
GPA	301	5	.22399	.52824	NS
	302	1	.57133	.57133	.01
	303	5	-.02863	.51025	NS
	304	4	.35820	.58773	NS
READ	301	2	-.16003	.46915	NS
	302	4	.05315	.77102	.01
	303	2	-.21119	.36295	NS
	304	2	-.35545	.54180	NS
AGE	301	3	-.13191	.49657	NS
	302	5	.05483	.77276	.05
	303	3	-.17875	.42301	NS
	304	1	.42301	.42301	NS

Cloze	%
301	31.5
302	34.0
303	26.9
304	49.8

for the anomalous results in the 200 series.

The tables are quite complex and may be approached from several different directions. First, the five place decimals in the columns marked Multiple R may be considered as indications of a relationship between the Cloze score and one of the three variables noted: GPA, reading attitude or age. Some of these relationships appear relatively strong, as .77102 on the reading question for form 302, or .65111 on GPA for form 102. However, as with the analysis of variance, a significance level is provided for each Multiple R calculated, and significance is once again crucial to the implications of the tables. In the majority of forms and for the majority of the variables the relationship indicated by the Multiple R figure is not significant (NS). This means that even though there is a relationship of the Cloze score to a variable in this sample, this relationship would very probably not hold for the population, and therefore, no statistically defensible conclusion can be drawn from the relationship, however strong it may appear.

Some of the Multiple R's are significant, however, at the .01 level of significance. This fact suggests another approach to Tables 4, 5, and 6. Perhaps there is a pattern of significant relationships either in the 200 series alone that would explain the anomalous Cloze scores, or in the 100 and 300 series but not in the 200 series that would explain the anomaly. In the 200 series, only two significant Multiple R relationships appear in the three variables under study. The Multiple R for the reading attitude question on form 201 is .55713, significant at .05, and the Multiple R for age on form 201 is .62063, also significant at .05. These relationships,

however, tell us nothing about the relationship of scores to these variables for 202, 203 and 204, or account in any way for the lack of significance in the changes in the Cloze scores for the modified forms in the 200 series.

Finally, there is the issue of the significance level found here. For some research, significance of .05 is considered acceptable for the purpose of making generalizations about the population. Perhaps this level should be accepted here. However, to do so might cause a skeptical reader to wonder why significance was set at .05 rather than the more trustworthy figure of .01. Also, even if .05 were taken as the level of significance, the data in Tables 4, 5, and 6 provide no clear pattern to explain the anomalous 200 series results.

Examining the tables from yet another direction, the 100 and 300 series regression results might set these series off from the 200 series, thereby accounting for the different results obtained. In the 100 series, only two significant Multiple R figures appear: .60050 for form 101 with GPA, and .65111 for form 102 with GPA. All the other Multiple R's, again, even though some of them appear to be strong relationships, are non-significant. In the 300 series, a different pattern of significance appears. The relationship of GPA, reading attitude and age to score is significant for form 302, but not for any of the other forms. This might tell us something about the Cloze scores in the 300 series, but it does not match, reflect or relate in any clear way to the significant Multiple R's in the 100 series nor does it distinguish the 100 and 300 series from the 200 series.

Two other sets of figures are provided in Tables 4, 5, and 6. One of these is the column marked Simple R. This column provides the simple relationship of the Cloze score to each variable considered alone. Here, among the three variables across all forms, the highest figure is .57133 for form 302, relating GPA to score. Since all the other Simple R's are below .50, indicating weak or non-existent relationships, nothing useful with respect to the 200 series results can be concluded here either.

The final column of these tables to be considered is the one marked step order. Step order indicates where in the step-wise regression each variable was entered on each form. Thus, the 1 under step order for form 103 with age indicates that age was taken first in the regression as the best predictor of Cloze score for form 103. This column might show that one variable was the best predictor of score in some pattern which would again account for the 200 series. In the 200 series, though, only two 1's appear--GPA was the best predictor in form 203 and age was the best predictor in form 202. Note that even if this were a pattern, the Multiple R's are non-significant. Looking at the 100 and 300 series in a similar fashion, the best predictors are few in number and show no clear pattern with respect to the 200 series.

The regression analysis done on the variables of GPA, reading attitude and age provides, therefore, no insight to the 200 series anomaly. Although interest and writing quality were also subject to regression, and will be discussed below, it is probably worth noting that neither variable is of any help. Because an explanation for the anomalous 200 series results was not found by regression, the

correlation ratio was used to analyze the remaining variables.

4. Correlation Ratio Results with Respect to the Anomalous Data

The remaining variables are all nominal scale variables (i.e. results fall into categories) and the question was whether there were relationships between these nominal variables and the Cloze scores which would provide some further insight to them. To correlate nominal and ordinal variables where their relationship is not linear, a correlation ratio or eta value is used.[4] Although this correlation is less satisfactory than regression because it can only relate a score to one variable at a time, it does allow an analysis of the remaining variables examined here. Further, by using the correlation ratio on all the nominal variables, we have a uniform comparison among them. The correlation ratios of Cloze scores to the pre-reading measure of interest (ranking) are presented in Table 7.

TABLE 7

CORRELATION RATIO OF CLOZE TO RANKING

Form	Cloze %	Political Science	Famous People	Sports	Women's Lib	Science
101	71.77	.692	.280	.446	.349	.525
102	56.73	.446	.185	.310	.355	.319
103	69.96	.438	.598	.555	.668	.552
104	81.25	.481	.452	.526	.438	.381
201	49.80	.876	.638	.708	.526	.674
202	43.50	.477	.203	.428	.439	.493
203	40.80	.317	.428	.544	.454	.407
204	40.10	.663	.390	.275	.682	.478
301	31.50	.638	.692	.652	.581	.535
302	34.00	.258	.569	.614	.452	.508
303	26.90	.320	.315	.403	.460	.428
304	49.80	.565	.581	.520	.667	.706

The correlation ratio is actually closer in nature to regression than to a more common variety of correlation like the Pearson correlation. The reason is that the correlation ratio is a measure of prediction--here, how well we can predict the Cloze score from the response on another variable--and not a measure of a direct relationship as is the case with Pearson. Again, like the regression analysis, the correlation ratio was used to try to find an explanation for the 200 series anomaly.

In Table 7, the correlation ratios or eta values are given for the relationship between the Cloze scores and the pre-reading interest measure called the ranking question. In this question, the subject was asked to rank five topics (political science, famous people, sports, women's liberation and science) in order of preference as topics to read about. All subjects numbered the five topics on the ranking question from one to five, "1" being used for the most preferred topic. The resulting data proved rather difficult to analyze, and as a result, the analysis should be viewed cautiously. Because the five topics represent five discrete categories, the correlation ratio indicates to what extent the ranking by a particular subject helps predict his Cloze score. It is probably easiest to think of this as a measure of correlation, although this is not precisely true.

In Table 7, as with the regression tables, the question is whether there is a pattern in the 200 series to account for the 200 series anomaly. Again, there are several ways of studying the table, and a question of criterion. A criterion of what constitutes a relationship must be set before the tables can be approached. In

this table, no significance is provided, so it is necessary to rely on the level of the eta value itself. One statistician[5] provides a guideline of $\pm$.40 to $\pm$.70 as a marked relationship, $\pm$.70 to $\pm$1.00 as a high relationship, and below $\pm$.40 as a low or negligible relationship for the Pearson correlation coefficient. Since this table is relying on the predictive eta value instead, it is best to take only values of $\pm$.70 and above as indicators of a strong predictive relationship.

We might, then, look for some connection between the ranking of topics in the 200 series and the scores. However, for the 200 series, the only strong relationship appears on form 201 between score and the topics political science and sports. If we look at the relationship of the ranking for the sports topic to the 200 series which was about sports, there is the strong relationship on the 201, but it does not hold for 202, 203, and 204. This provides no apparent insight to the 200 series.

To look at the 100 and 300 series, no strong relationships of any of the topics to Cloze score appear for any form on the 100 series, and in the 300 series the only strong relationship appears on form 304 with the science topic. If we look at the relationship of the ranking of the topic which is the subject of the series, women's liberation for the 100 series, and political science for the 300 series, no clear pattern of strong relationships appears.

If a weaker relationship were considered acceptable, allowing $\pm$.40 and above as the criterion, for example, no pattern emerges. There would in this case be some evidence for a relationship of Cloze score to the ranking of the political science topic in the 100 series,

but then an almost identical pattern appears on the science topic for the 200 series.

Treated in these various ways, Table 7 appears to yield no strong relationship of Cloze score to the ranking of the five topics in the pre-reading interest question. Such relationships as do appear in Table 7 do not provide any further explanation of the 200 series anomaly. The raw results of the ranking question suggest an explanation of the 200 series when considered together with the post-reading interest rating. No statistical analysis confirms this possibility, however, so discussion of it is postponed to chapter 4.

Table 8 presents the results of the correlation ratio analysis of the remaining variables of sex, major, degree sought (B.A., B.S., B.F.A. and so on), leisure time activity (leisure), the type of language(s) besides English studied by the subject (Romance, Germanic, and so on, listed as lang. type) and the number of languages besides English studied by the subject (lang. numb.). Table 8 must be analyzed using the same strategies as for Table 7. In Table 8, some distinction in the 200 series or some characteristic(s) shared by the 100 and 300 series might account for the 200 series results.

To study the table, the eta values for each variable may be read down, to compare the 200 series to the 100 and 300 series. If the criterion for a relationship is set at $\pm$.70 and above, only one relationship appears--between Cloze score and choice of leisure time activity on form 104. Because of this, it might be more profitable to lower the criterion to $\pm$.40 and above. Even if the criterion is lowered, however, no explanatory pattern emerges. For sex, only one relationship to score appears, on form 204. For major, virtually all

TABLE 8

CORRELATION RATIO OF CLOZE TO SELECTED VARIABLES

Form	Cloze %	Sex	Major	Degree	Leisure	Lang. Type	Lang. Numb.
101	71.77	.138	.609	.665	.390	.143	.070
102	56.73	.099	.343	.672	.589	.362	.072
103	69.96	.374	.545	.329	.228	.489	.176
104	81.25	.132	.451	.320	.716	.332	.119
201	49.80	.258	.513	.096	.451	.299	.317
202	43.50	.298	.563	.217	.678	.469	.225
203	40.80	.019	.521	.255	.490	.564	.005
204	40.10	.475	.473	.131	.348	.419	.006
301	31.50	.115	.583	.432	.638	.184	.000
302	34.00	.344	.644	.166	.470	.208	.036
303	26.90	.154	.619	.315	.423	.177	.058
304	49.80	.138	.543	.169	.495	.662	.080

forms show a relationship to score. These weaker relationships disappear as one reads to the right, so that the variables degree, leisure time activity and type and number of languages studied have little or no predictive value with respect to the Cloze scores.

Alternatively, Table 8 might be read across, looking for some variable which distinguishes the 200 series from the others. Even with a lowered criterion, only major and leisure time activity show some relationship to score in the 200 series, and then by going back to the vertical orientation, it is clear that these variables also show some relationship in both the 100 and 300 series for major and the 300 series for leisure time activity. The relationships that do appear, therefore, fail to distinguish, explain or clarify the anomalous Cloze scores. That is, Table 8, like Table 7, appears to show no strong relationships which would explain or account for the

200 series results.

Thus, the correlation ratios or eta values calculated for all the variables not involved in the regression provide no explanation or insight into the 200 series. There appear to be no statistically valid explanations for the 200 series results, given the hypotheses and design used here. Several possible explanations can be derived from the raw data on several of these variables, but none of these speculations can be confirmed statistically. For this reason, discussion of these possibilities is postponed until chapter 4.

5. Regression Results with Respect to the Corollary Hypotheses

In the discussion of Tables 4, 5, and 6, the results of the regression analysis on the post-reading measures of interest and writing quality were not discussed, since they provide data for the corollary hypotheses of the study that increasing redundancy will increase interest and improve subjective evaluations of writing quality. The results of the regression on the interest and quality variables can be analyzed in the same terms as the other regression results, with, unfortunately, essentially the same outcome.

Starting to the far right in Tables 4, 5, and 6, there are a number of promising multiple R figures for interest and quality, but again, the significance levels tell a different story. If significance is set at the preferable level of .01, significant relationships appear only on form 302 between interest and score and between quality and score. If .05 is accepted as the level of significance, significant relationships appear between interest and score on forms 102 and 203, between quality and score on forms 102

and 201. The Simple R figures are similarly not helpful, and step order also provides no insight.

In view of the Cloze scores, especially where there was significant improvement in the means in the 100 and 300 series, it was expected that the interest and quality measures would also show an improvement. However, the regression analysis shows no consistent significant relationships between Cloze scores and the variables of interest and quality, nor any improvement or lowering in the ratings as the redundancy increases. The raw data on the post-reading interest question are suggestive of several possible conclusions, none of which can be confirmed statistically. These possibilities are pursued in chapter 4.

6. Summary

The results of a statistical analysis of all the data gathered in the research have been reported here. The major result of the research is the analysis of variance on the Cloze test scores which shows that scores on the 04's are significantly higher than those on the 01's in two of the three series, and there is no significant difference between the 01 and 04 in the third case. Correlation of the other variables to the Cloze scores by regression and the correlation ratio was attempted to try to gain further insights into the Cloze test results and the two corollary hypotheses. The correlations yield no consistent significant insights to either the Cloze scores or the corollary hypotheses. Some of the raw data are suggestive of possible relationships between variables and Cloze scores. These possibilities are pursued in chapter 4.[6]

FOOTNOTES: CHAPTER 3

[1]William Wiersma, Research Methods in Education, 2nd ed. (Itasca, Illinois: F. E. Peacock Publishers, 1975), pp. 189-91.

[2]"Regression," in Statistical Package for the Social Sciences, ed. Nie, et al. (New York: McGraw-Hill Book Company, 1972).

[3]I spoke to the Registrar at Wayne State University, and was refused access to university computer files containing that information. The income data that I did get was probably inaccurate, since some subjects quoted figures of $5,000 and others quoted figures of $50,000. Finally, had we used some method of averaging the present data to fill in the absent data, this might have resulted in a "zero divide" situation in the computer, which would have caused the machine to throw out the job.

[4]Wiersma, p. 302f. The eta value and its calculation is discussed fully in John Mueller, Karl Schuessler and Herbert Costner, Statistical Reasoning in Sociology, 2nd ed. (New York: Houghton Mifflin Company, 1970), pp. 325-33.

[5]Wiersma, p. 311.

[6]For assistance in calculating the Pearson correlation coefficient for the reliability study, and for general assistance in understanding statistics, I am indebted to Mr. William Horning of Hastings, Michigan, and to my husband, Arthur Horning.

CHAPTER 4

DISCUSSION, SUMMARY AND CONCLUSIONS

1. Introduction

One major hypothesis and two corollary hypotheses were investigated. The major hypothesis is that increasing redundancy increases comprehensibility. The corollary hypotheses are that increasing redundancy increases interest in a passage and that increasing redundancy improves the subjective evaluation of the writing quality of a text. The conclusions on the major hypothesis are based on the analysis of the mean Cloze scores on each form. The conclusions on the corollary hypotheses are based on the regression analysis of the post-reading ratings of interest and writing quality and on the pre-reading ranking of five reading topics, analyzed by the correlation ratio. The conclusions for each hypothesis are treated in separate sections of this chapter. The conclusions on the major hypothesis are followed by a discussion of some of the raw data from the research which provide further support for some of the conclusions. The conclusions are followed by a brief sketch of several avenues for further research suggested directly or indirectly by the data.

2. Conclusions on the Major Hypothesis

The results of the Cloze tests and an analysis of them show a significantly higher Cloze score on the 04 forms (those with both

types of redundancy added) as compared to the 01's on two of the three topics, and no significant difference in score on the 04 as compared to the 01 on the third topic. The results also show no consistent significant differences between the 02 version (with increased syntactic redundancy) and the 03 version (with increased semantic redundancy) as compared to the 01's (originals).

Based on these results, the major hypothesis appears to be confirmed in a limited and qualified sense. That is, increased redundancy, when redundancy is increased both syntactically and semantically, seems to increase comprehensibility. As noted above, in one case (the 200 series which was the passage on sports), increasing redundancy, even of both types, made no significant difference in the scores. All other variables in the study were analyzed using either regression or the correlation ratio in an attempt to account for the anomaly of the 200 series. No consistent, significant relationships were found among the variables, or between any of the variables and the Cloze scores, to account for the 200 series. However, several observations about the mean scores, the linguistic qualities of redundancy, and reader background do provide further insight into the results.

The analysis of variance was explained previously as, in part, a comparison of the effect of the presence of a variable to its absence. If the means are combined to reflect the treatment of the analysis of variance on each variable alone, the results appear as in Table 9. Note that, from the analysis of variance, semantic redundancy alone significantly (*) improves the means for the 100 series and syntactic redundancy alone improves the means for the 300

TABLE 9

CLOZE MEANS COMBINED AS TREATED BY ANALYSIS OF VARIANCE

Without Syntactic Redundancy		With Syntactic Redundancy		Without Semantic Redundancy		With Semantic Redundancy	
Forms	Mean	Forms	Mean	Forms	Mean	Forms	Mean
101 103	32.65	102 104	33.40	101 102	31.13*	103 104	34.93*
201 203	22.65	202 204	20.90	201 202	23.33	203 204	20.23
301 303	14.60*	302 304	20.95*	301 302	16.38	303 304	19.20

series, and this increase is also significant. Note also, however, that the combined means do improve for syntactic redundancy alone in the 100 series and semantic redundancy alone in the 300 series, albeit non-significantly. The 200 series presents a contrast in that the combined means go down in the case of either type of redundancy and the mean on 204 is also lower than that on 201. These drops are non-significant, but are a contrast to the 100 series and the 300 series, where adding redundancy improves the means to some extent in all cases. Although the drop in the means in the 200 series is non-significant, that is, neither supporting or disconfirming the major hypothesis, the drop shows precisely this contrast between the 200 series and the other two sets of passages.

Further examination of the means shows another difference. In the 100 series, the mean scores are relatively high in terms of percentage correct. A Cloze percentage of over 50% is a relatively high score, and the average Cloze percentage for all of the 100 series

is 69.93%. In the 300 series, the mean percentage is relatively low. Here the average Cloze percentage for all of the 300 series is 35.55%. By contrast, the 200 series falls closer to 50% than either of the other two, with an average Cloze percentage of 43.55%. In view of these facts, the 100 and 300 series might be seen as extremes on a scale of difficulty, with the 100 series representing a relatively easy text, and the 300 series a relatively difficult text.

If these observations about difficulty are true, then the difficulty of a text is apparently not directly related to sentence length and vocabulary level. These superficial predictors of difficulty were held constant for all versions of all passages, and yet, there is a substantial difference in the Cloze scores on these passages. Difficulty of a text is evidently significantly related to linguistic redundancy. The effect of experiential redundancy, i.e. the reader's prior knowledge, must also play a role and deserves further study. Finally, difficulty is apparently significantly influenced by reader interest, along with a number of other factors not studied here, such as motivation.

Redundancy seems to improve comprehensibility, then, at the extremes of ease or difficulty, but not in the middle. In part, this pattern may exist because when Cloze scores hover around 50%, the material is thought to be at an appropriate readability level for the reader.[1] When this level appears as a mean Cloze score, redundancy appears to distract the reader from comprehending, rather than to aid him.

One of the subjective impressions one gets from looking at the actual fill-in items on the Cloze tests is that where the subject

filled a blank, the word put in is almost always syntactically acceptable with respect to the original item, even though not always correct semantically. This suggests an explanation for the significance pattern of the Cloze results. When the material is easy, as in the 100 series, the reader can be more concerned with the details of meaning, so that syntactic redundancy improves the scores only slightly, but semantic redundancy improves comprehensibility significantly. When the material is relatively difficult, as in the 300 series, making the syntactic structure more predictable by adding redundancy improves comprehensibility significantly. However, semantic redundancy--repetition and detailed explanation of meaning--only improves scores slightly.

These observations point to an important implication of the results: not only does adding redundancy increase comprehensibility in general, but also, once the Cloze score on a text is known for a particular group of readers, it appears possible to decide both whether adding redundancy will make a significant difference, and also what type of redundancy will yield the most significant improvement in comprehensibility. Thus, an original text showing a Cloze score of 70% or above for a group of readers (like the 100 series) can have its comprehensibility increased significantly by adding semantic redundancy. A text with a Cloze score of 30% or below for a group of readers (like the 300 series) can have its comprehensibility increased significantly by adding syntactic redundancy, and for one where the Cloze score is at about 50%, increasing redundancy will probably not help at all.

In addition to providing a very practical insight to the use of redundancy in texts, these observations also help to illustrate the fact that the definition of redundancy in operational terms on a syntactic and semantic dimension is a viable definition. The term viable is quite appropriate here: the definition is not only workable, but also is capable of further growth and development. Since an operational definition of redundancy has been lacking until now, and since the development of this definition was one of the major goals of this project, the confirmation of the definition, albeit in a limited sense, is a major result of this project. The definition will undoubtedly be modified and improved upon in further research.

The fact that the 200 series text was at the right level for the subjects in this study does not, by itself, fully account for the fact that added redundancy made no significant difference in the comprehensibility of the text, or that in a non-significant way added redundancy seems to interfere with comprehensibility. The raw data from the two measures of interest used in the study suggest another way in which the 200 series differs from the other texts. Although the points made in the discussion which follows cannot be confirmed statistically, when considered together with the observations above, they represent mounting evidence accounting for the anomaly of the 200 series Cloze scores.

From a subjective point of view, reader interest would seem to be a key factor in comprehensibility. Most people find technical material or scientific studies difficult to comprehend unless they are specialists, whereas a compelling mystery is quite easy to comprehend. Speed of reading is an obvious indicator of this

difference: most people wish they could read technical material as quickly and with the same high comprehension they get for novels, newspapers, and the like. High interest, however, can offset difficulty of material, so that if the reader is very interested, he will probably become familair with the subject quickly and make steady progress through even a very difficult text. Low interest, however, has the opposite effect, with some readers falling asleep over difficult texts. It appears, in the raw data on interest, that low interest is the case in the 100 and 300 series, but not in the 200 series.

The raw data on interest come first from the pre-reading ranking question, where the subject was asked to rank five topics in order of his preference for them as reading subjects. The topics in the list included women's liberation (the topic of the 100 series), sports (200 series) and political science (300 series). Two other "dummy" topics were also put on the list: famous people and science. Each subject numbered the topics starting with "1" for the topic he most liked to read about. In Table 10, the rankings by all subjects for the topics of the passages they read are tabulated. That is, all the responses to the women's liberation topic for the subjects who read the 100 series are given in the table, and so on. For form 101, one subject ranked women's liberation first, three subjects ranked it second, five subjects ranked it third, and so on. Again, a ranking of "1" indicates high preference for, or interest in, the topic, whereas a ranking of "5" indicates low interest. Only the ranking of the topic from the list which is also the topic of the passage read is given here. That is, for the 100 series, the rankings on women's liberation are given, for the 200 series, the rankings on

TABLE 10

TABULATION OF RANKINGS FOR THE TOPIC OF EACH PASSAGE

Form	Cloze %	1	2	3	4	5	
101	71.77	1	3	5	4	7	
102	56.73	2	4	6	2	5	Ranking of Topic 4
103	69.96	0	1	7	6	5	Women's Liberation
104	81.25	2	4	5	3	6	
201	49.80	6	1	5	2	5	
202	43.50	4	9	5	0	2	Ranking of Topic 3
203	40.80	5	6	5	1	3	Sports
204	40.10	6	4	4	2	1	
301	31.50	3	2	4	6	3	
302	34.00	1	0	8	3	6	Ranking of Topic 1
303	26.90	0	2	4	5	9	Political Science
304	49.80	2	3	4	6	4	

sports as a topic are given, and for the 300 series, the rankings on political science are given.

In comparing the number of high interest rankings given to the women's liberation topic by the 100 series subjects to the number of high interest rankings given to sports by the 200 series subjects, and those given to political science by the 300 series subjects, the 200 series subjects once again stand out. If rankings of "1" and "2" are combined for each series as high interest, and "4" and "5" are combined as low interest, the contrast is even more clear. Subjects in the 100 and 300 series were not as interested in the topics they read about as were the subjects in the 200 series.

Although this observation cannot be statistically confirmed, the raw data on the post-reading rating of subject interest in the passage read indicate a similar contrast. Table 11 tabulates the

results of the post-reading interest measure, which asked: "Despite your ranking of the topics above, how interesting did you find this passage?"

TABLE 11

TABULATION OF POST-READING INTEREST QUESTION RESULTS

Form	Cloze %	1	2	3	4	5
101	71.77	1	8	5	3	3
102	56.73	1	5	7	5	2
103	69.96	1	9	6	2	2
104	81.25	1	4	6	5	4
201	49.80	3	4	4	5	4
202	43.50	0	7	4	5	4
203	40.80	1	2	7	7	3
204	40.10	0	7	6	6	1
301	31.50	0	8	5	3	4
302	34.00	0	1	7	2	10
303	26.90	0	3	3	5	8
304	49.80	1	2	9	4	3

Key:
1 = very interesting
5 = very boring

The subject circled one of five choices ranging from very interesting (1) to very boring (5). The 200 series again stands out, particularly if 1's and 2's are combined as high ratings of interest, and 4's and 5's are combined as low ratings. The 200 series subjects appear more interested in their passage on sports than the other subjects were in their passages. Like the observations about Table 10, these observations cannot be confirmed statistically, but taken together with the Cloze scores, they suggest why increasing redundancy made no significant difference to the Cloze scores in the 200 series.

The relationship between interest and the Cloze results seems to be that increasing redundancy significantly improves comprehensibility in the presence of lower interest as in the 100 and 300 series, but has no significant effect in the presence of high interest. Common sense seems to support this relationship, as we have observed that readers will wade through even the most complex and prosaic texts if they are interested. In such a case, adding redundancy seems not to make a significant difference, except perhaps to distract the reader slightly. The implication of this observation for much written material is important. Where low interest is assumed or expected, a writer might artificially increase redundancy to improve comprehensibility. This sort of text modification could make a substantial difference in technical materials and also in pedagogical writing.

One further explanation for the disparity of the 200 series is provided by the raw data from the leisure time activity question, and relates to the issue of the role of redundancy as it interacts with the reader's background. Clearly, some redundancy is created in text by how much a reader knows about a topic beforehand. If readers know a lot about a topic, reading material on that topic is likely to be at least somewhat familiar to them and therefore, somewhat redundant because of their background. Thus, what the reader brings to a text may have some effect on how comprehensible the text is and to what extent it is already redundant. Table 12 presents a tabulation of leisure time activity preferences by category. The main point here is that if each column is totalled, giving a composite picture of all subjects in the study, more subjects indicated an activity in the sports category than in any other.

TABLE 12

TABULATION OF LEISURE TIME ACTIVITY PREFERENCES BY CATEGORY

Form	Cloze %	1	2	3	4	5	6	7
101	71.77	5	1	4	0	0	1	5
102	56.73	9	3	2	0	1	2	3
103	69.96	5	3	5	1	0	3	2
104	81.25	4	2	5	1	2	3	3
201	49.80	7	0	9	2	0	0	1
202	43.50	3	1	5	3	1	3	2
203	40.80	7	2	1	1	0	6	3
204	40.10	10	3	1	0	2	4	0
301	31.50	7	0	1	0	3	5	2
302	34.00	5	2	5	1	1	4	1
303	26.90	7	2	4	2	0	2	3
304	49.80	7	0	2	3	2	1	3
Totals		76	19	44	14	12	34	28

Key:
1 = sports
2 = socializing
3 = fine arts
4 = hobbies
5 = tv
6 = misc.
7 = reading

Like the other observations from the raw data, this one cannot be statistically confirmed, but it suggests something more about the 200 series. Sports was an area of preference for many of the subjects, and those reading the 200 series may therefore have brought more background about sports to the reading task, making the passages, in all forms, familiar without linguistic modification. Thus, adding redundancy seems to have had little or no impact, perhaps because the text was already familiar to the subjects, by virtue of their interest in, or prior knowledge of, the sports topic.

With respect to the major hypothesis, then, it appears that increasing redundancy does increase comprehensibility in at least two of the three cases studied here. The linguistic qualities of redundancy and some of the reasons why it had no effect on comprehensibility in the third case studied have been described and discussed. Although none of these observations can be statistically confirmed and each requires much further study, taken together they provide a multi-faceted explanation for the 200 series results. The results also show that the definition of redundancy devised for this study is a viable definition and that practically speaking, the two types of redundancy significantly increase comprehensibility under highly specified circumstances.

3. Conclusions on the Corollary Hypotheses

One of the corollary hypotheses of the study is that increasing redundancy increases interest in a passage of text. Two measures of interest provided data on this hypothesis: the pre-reading ranking question which asked the subjects to indicate by rank order their interest in five topics, and the post-reading interest rating of the passage read. A correlation ratio analysis was used on the ranking question and regression was used to analyze the responses to the rating question. Both analyses show that at best only weak correlations appear between the Cloze scores and the two measures of interest, and at worst, there is no correlation at all. Even where there is a correlation between Cloze score and interest, such correlations are generally not consistently significant.

Based on these results, the corollary hypothesis on interest is not confirmed by the data. Part of the problem with the interest hypothesis is the way in which the data on interest were collected. The pre-reading ranking question was probably a good idea in theory, but difficult to use in practice. The ranking question introduced two dummy topics: science and famous people and the latter was an especially popular choice. These topics may have skewed the results. The data as collected were hard to handle because the ordering of topics by each subject could not be fully considered in the correlation ratio, and the rankings also could not be analyzed by regression or any other more standard correlation procedure, because of the nominal data generated. Finally, it is not clear that this ranking question, even considered together with the post-reading interest measure, represents a truly accurate measure of reader interest.

Actually, the raw data on interest discussed above suggest something quite different about the interest factor. It is certainly clear that interest is important, but in addition, interest seems to have some effect on redundancy in a text. For this reason, the terms of the corollary hypothesis may be inverted order. A more appropriate statement may be: increased interest diminishes the effect of increased redundancy. This is the conclusion the raw data suggest. Although this point is not statistically confirmed, it is supported by the analysis of the major hypothesis results discussed previously. The role of interest in reading presents a complex problem as there is no widely accepted, valid, reliable measure of interest, and, as yet no one has looked at interest with respect to the impact of redundancy. If interest has a controlling effect on the impact of

increased redundancy, it warrants very thorough further study.

The second corollary hypothesis on writing quality is that increasing redundancy improves the subjective evaluation of the writing quality of the passage. A regression analysis was done on the post-reading quality question, which asked the subject to evaluate the quality of the writing in the passage read. The analysis showed no consistent, significant correlations between the quality ratings and the Cloze scores.

Based on these results, the corollary hypothesis on quality is not confirmed by the data. Like the interest hypothesis, this hypothesis presented some problems in the way the data were collected. The subjects were asked to rate the passage in terms of clarity, organization and style. Despite the fact that these are relatively specific criteria, some subjects may not have been able to apply them. (A recent survey done in the English Department at Wayne State University showed that even the instructors of the English 0150 Freshman Composition course do not agree on the nature and meaning of these criteria.)

In the case of both corollary hypotheses, even if any conclusions could be drawn, they might be invalidated by the fact that only one direct question was asked for each issue. Since it was not clear what importance either issue might have, this method seemed reasonable at the outset. Although this strategy for data collection on the corollary hypotheses weakens the study as a whole, the raw data on the interest question seem to suggest that the interest factor should be very thoroughly examined.

4. Directions for Further Study

The amount of discussion devoted to the raw data in this chapter suggests that several follow-up studies are needed for two reasons. First, several passages at different Cloze levels for the subjects should be tested to clarify the 200 series results. Second, follow-up studies are needed to investigate the effect of interest and reader background on redundancy and comprehensibility. Additional work needs to be done to develop a valid, reliable measure of reader interest which will provide ordinal data on interest. Such ordinal data would make possible clear Pearson correlations of Cloze scores to interest scores. Also, an ordinal measure of writing quality would yield comparable data on reader judgment and perhaps make it possible to confirm the hypothesis on writing quality. However, teachers of writing have been looking for a measure of this kind for some time without success, and it is not clear that this would be a fruitful endeavor.

Although the raw data provide multiple avenues for further investigation, several other possibilities have emerged during the study. One such possibility involves text sophistication. Although the Dale-Chall readability formula was used to control the sentence length and vocabulary level of the material, perhaps the sophistication of the passages vis a vis the subjects was a factor. One way of testing this might be to try the same passages on older or younger subjects to see if their sophistication or lack of it with respect to the text might make a difference. Another possibility involves the style of the passages. Perhaps the 100 and 300 series texts are "highly abstruse," whereas the 200 series text is not. Such an

investigation would have to find or develop a style index to measure a factor like "abstruseness." Although the interest factor has already been mentioned as a subject for further study, a further possibility would involve checking on the high score/low interest relationship. Here, subjects would have to be pre-selected in two groups--one with high interest and one with low interest in the topic of the passage. Some preliminary research would be in order in this case to develop a valid, interval scale index of reader interest. Such a scale may prove extremely difficult to develop and validate.

Because of problems discussed previously, this study disregarded the socio-economic and educational background of the subjects. This factor may also prove important to the issue of redundancy and could be important to developmental as well as proficient adult reading. Finally, far in the future, when the effect of increased redundancy on comprehensibility of English text for native speakers of English is clear, researchers might turn to students of English as a second language (ESL) and their problems learning to read English proficiently. Although the ESL problems which led to this study originally cannot be solved, at least some of the paths for further study are marked, and may ultimately lead to solutions to the problems in ESL.

5. Summary

The present study has, of course, done more than simply mark paths for further study, though as a groundwork study, the marking of paths is important. The research has several larger potential implications for the field of reading. These implications are noted

as potential because these results have yet to be confirmed and because so much additional work remains to be done. The definition of redundancy devised for this study and confirmed in a qualified sense in the major hypothesis implies that redundancy can play an important role in making the meaning more available in reading. It is also now clear that redundancy can be easily increased in texts, and that increasing redundancy is desirable. Finally, this area warrants further study and is likely to yield valuable insights to increasing the comprehensibility of text. The psycholinguists on whose work this study rests define reading as getting meaning from print. This research has shown that by adding redundancy to a text, we may be able to help a reader achieve this goal more easily.

FOOTNOTES: CHAPTER 4

[1]John Bormuth, "Comparable Cloze and Multiple Choice . . . Scores," Journal of Reading, 10 (1967), 291-99.

APPENDIX 1

RESULTS OF THE PILOT STUDIES

APPENDIX 1

RESULTS OF THE PILOT STUDIES

TABLE 13

RESULTS OF THE FIRST PILOT

Form	N	Raw Mean/No. of Blanks	Pct.
100 Series: Male Dominance			
101	14	16.93/50	33.86
102	13	17.23/49	35.16
103	12	17.50/49	35.71
104	18	23.05/48	48.02
200 Series: Sports			
201	13	23.00/50	46.00
202	11	23.54/49	48.04
203	18	21.78/48	45.38
204	9	19.78/51	38.78
300 Series: Government			
301	13	17.31/50	34.62
302	13	17.77/50	35.54
303	14	17.79/50	35.58
304	9	14.44/49	29.47

TABLE 14

RESULTS OF THE SECOND PILOT

Form	N	Pilot 2 %s	Pilot 1 %s
101	-	not run	33.86
102	-	not run	35.16
103	-	not run	35.71
104	15	39.17	48.02
201	17	46.94	46.00
202	8	37.72	48.04
203	16	45.12	45.38
204	13	43.38	38.78
301	16	32.12	34.62
302	14	28.00	35.54
303	12	31.16	35.58
304	14	30.61	29.47

APPENDIX 2

DIRECTIONS TO TEST ADMINISTRATORS

APPENDIX 2

DIRECTIONS TO TEST ADMINISTRATORS

Directions for experiment

1. Thank you, again, and please thank your students, for me, for their help.

2. Ask the students to answer the questions on the top half of side one. Please tell the students the experiment will not have any impact on their grades or careers.

3. Tell the students I am measuring the readability of this material for college students.

4. Tell them they have 20 minutes to do the exercise on side 2.

5. Give your class 20 minutes to work as directed on side 2. Do NOT give them more than 20 minutes, even if some students do not finish the exercise in the time allowed.

6. At the end of the 20 minutes, ask the students to turn back to side 1 and answer the two questions at the bottom of the page. Try to be sure everyone answers these questions.

7. Collect the materials and place them in my mailbox at your very earliest convenience.

8. Many thanks, again, for your help.

NOTE: Not all students will read the same passage, in case you try to discuss it after the exercise.

APPENDIX 3

MATERIALS

APPENDIX 3

MATERIALS

TO THE STUDENT: Please respond to the following items:

Your instructor's name: _____

Sex: male _____ female _____

Age: _____

Your major, or, if insure, subject which interests you most now: _____

Are you a U. S. citizen? (If not, indicate native country) _____

Year and term in school (e.g. senior, second term): _____

What language(s) do you speak at home? _____

Overall grade point average: _____

What degree are you working toward? (B.A., B.S., etc.) _____

What is your parents' average yearly income? _____

What is your favorite leisure time activity? _____

Which languages have you studied besides English? _____

How often do you read for pleasure (newspapers, books, etc.)?
Circle one: every day a few times a week about once a week very rarely

Which of the following topics would you most like to read about? Rank them according to your preference, using 1 for the topic you like most:

_____political science

_____famous people

_____sports

_____women's liberation

_____science

QUESTIONS:

Directions: When you are told to do so, please answer the following questions by circling one answer for each question.

1. Despite your ranking of topics above, how interesting did you find this passage?
   a. very interesting
   b. moderately interesting
   c. neither interesting nor boring
   d. moderately boring
   e. very boring

2. In terms of clarity, organization and style, how well written do you think this passage is?
   a. very well written
   b. moderately well written
   c. neither good nor bad
   d. moderately poorly written
   e. very poorly written

Form 101

Directions: Please read the following passage through as many times as you need to, filling in the blanks with the words you think have been left out. Some words may be used more than once, and words already in the passage may be used. Use only one word in each blank. Try to fill in every blank.

So far as the history of the origin of masculine dominance is concerned, we must call attention to the fact that this is a phenomenon which does not occur as a natural thing. This is indicated by the numerous laws which are necessary to guarantee this domination to men. It is also an indication that previous to the legal enforcement of masculine domination there must have been other epochs in which the masculine privilege was not nearly so certain. History proves that such epochs actually existed in the days of the matriarchate, the age in which it was the mother, the woman, who played the important role in life, particularly so far as the child was concerned. At that time each man in the clan was in duty bound to respect the honored position of the mother. Certain customs and usages are still colored by this ancient institution, as for instance, the introduction of all strange men to a child with the title "uncle" or "cousin". A terrific battle must have preceded the transition from matriarchate to masculine domination. Men who like to believe that their privileges and prerogatives are determined by nature will be surprised to learn that men did not possess these prerogatives from the beginning, but had to fight for them. The triumph of man was simultaneous with the subjugation of women, and it is especially the evidence in the development of the law which bears witness to this long process of subjugation. Masculine dominance occurred chiefly as a result of constant battles between primitive peoples, during the course of

which man assumed the more prominent role as warrior, and finally used his newly won superiority in order to retain the leadership for himself and for his own ends.

Form 102

We must call attention to the fact that so far as the history of the origin of masculine dominance is concerned, this phenomenon of masculine dominance does not occur as a natural thing. The fact that masculine dominance is not natural is indicated by the numerous laws which are necessary to guarantee this domination to men. Previous to the legal enforcement of masculine domination, there must have been other epochs in which the masculine privilege was not nearly so certain, and this fact also indicates that masculine dominance is not natural. History proves that such epochs actually existed in the days of the matriarchate, the age in which the mother, the woman played the important role in life, particularly so far as the child was concerned. Each man in the clan was in duty bound, at that time, to respect the honored position of the mother. Certain customs and usages are still colored by this ancient institution, as for instance, the introduction of all strange men to a child with the title "uncle" or "cousin". A terrific battle must have preceded the transition from matriarchate to masculine domination. Men may like to believe that their privileges and prerogatives are determined by nature, but they will be surprised to learn that men did not possess these prerogatives from the beginning, but had to fight for them. The triumph of man was simultaneous with the subjugation of women, and the evidence in the development of the law especially bears witness to this long process of subjugation. Masculine dominance occurred chiefly as a result of constant battles between primitive peoples, during the course of which man assumed the more prominent role as

warrior, and finally used his newly won superiority in order to retain the leadership for himself.

Form 103

So far as the history of the origin of masculine dominance is concerned, we must call attention to the fact that this is a phenomenon which does not occur as a natural thing. This is indicated by the numerous laws which are necessary to guarantee this domination to men, such as property rights and divorce laws. It is also an indication that previous to the legal enforcement of masculine domination there must have been other epochs in which the masculine privilege was not nearly so certain. History proves that such epochs actually existed in the days of the matriarchate, the age in which it was the mother, the woman, who played the important role in life, particularly so far as the child was concerned. At that time, each man in the clan was in duty bound to respect the honored position of the mother. Certain customs and usages are still colored by this ancient institution, as for instance, the introduction of all strange men to a child with the title "uncle" or "cousin". A terrific battle must have preceded the transition from matriarchate to masculine domination. Men who like to believe that their privileges and prerogatives, for example, their authority in the family, are determined by nature will be surprised to learn that men did not possess these prerogatives from the beginning, but had to fight for them. The triumph of man, that is, his rise to a position of power, was simultaneous with the subjugation of women, and it is especially the evidence in the development of the law which bears witness to this long process of subjugation. Masculine dominance occurred chiefly as a result of constant battles between primitive peoples, during the course of which man assumed the more prominent role as warrior.

Form 104

We must call attention to the fact that so far as the history of the origin of masculine dominance is concerned, this phenomenon of masculine dominance does not occur as a natural thing. The fact that masculine dominance is not natural is indicated by the numerous laws which are necessary to guarantee this domination to men, such as property rights and divorce laws. Previous to the legal enforcement of masculine domination, there must have been other epochs in which the masculine privilege was not nearly so certain, and this fact also indicates that masculine dominance is not natural. History proves that such epochs actually existed in the days of the matriarchate, the age in which the mother, the woman played the important role in life, particularly so far as the child was concerned. Each man in the clan was in duty bound, at that time, to respect the honored position of the mother. Certain customs and usages are still colored by this ancient institution, as for instance, the introduction of all strange men to a child with the title "uncle" or "cousin". A terrific battle must have preceded the transition from matriarchate to masculine domination. Men may like to believe that their privileges and prerogatives, for example their authority in the family, are determined by nature, but they will be surprised to learn that men did not possess these prerogatives from the beginning, but had to fight for them. The triumph of man, that is, his rise to a position of power, was simultaneous with the subjugation of women, and the evidence in the development of the law especially bears witness to this long process of subjugation. Masculine dominance occurred chiefly as a result of constant battles between primitive peoples.

Form 201

American newspapers created, nurtured and shaped mass spectator sports between the end of the Civil War and World War II. Since World War II, radio and television have become more important in spreading the interest and determining the form of these activities. In no other culture have the commonplace aspects of popular games so permeated daily life, even for those who pay no direct attention. Sports terminology and thinking habits have become an inseparable, often unrecognized, part of our language and even our philosophy. In man-hours devoted to sports activities, in dollars generated by direct and allied sports business, in sheer amount of recreational attention devoted to playing, watching, reading about and listening to organized sports events, American sports have a hold on American civilization unmatched in any other time or place.

Spectator interest rests on three pillars: intimate knowledge (or at least the illusion of it) of the nature of the game and the identity of the participants, some degree of emotional alliance to one side or the other, and willingness to spend money for this sort of entertainment. Without the first, mental involvement is not possible; without the second, the first quickly becomes meaningless. But without both, the motivation for the crucial third element can't be maintained, and without the commercial element the staging of sufficiently skilled and significant contests would not be possible.

The growth of television and radio, however, has meant an important difference to the first element, the spectator's knowledge. Television has changed this knowledge in two directions. On the one hand it is immeasurably more informative than newspaper and magazine

stories ever were: the close-up, the instant replay, the right camera angle, slow motion. Furthermore, television (and even radio) carries with it instantaneous commentary on what is happening, complete with statistical and personal sidelights.

Form 202

American newspapers created, nurtured and shaped mass spectator sports between the end of the Civil War and World War II. Radio and television have become more important in spreading interest and determining the form of these activities since World War II. The commonplace aspects of popular games have permeated daily life in our culture as in no other, even for those who pay no direct attention. Sports terminology and thinking habits have become an inseparable, often unrecognized, part of our language and even our philosophy. American sports have a hold on American civilization unmatched in any other time or place, in man-hours devoted to sports activity, in dollars generated by direct and allied sports businesses, and in sheer amount of recreational attention devoted to playing, watching, reading about and listening to organized sports events.

Spectator interest rests on three pillars: intimate knowledge (or at least the illusion of knowledge) of the nature of the game and the identity of the participants, some degree of emotional alliance to one side or the other, and willingness to spend money for this sort of entertainment. Mental involvement is not possible without the knowledge; the knowledge becomes meaningless quickly without emotional alliance to a side. The motivation for the crucial commerical element can't be maintained without knowledge and emotional alliance, and without the commercial element, the staging of sufficiently skilled and significant contests would not be possible.

The growth of television and radio has meant an important difference to the spectator's knowledge, however. Television has changed this knowledge in two directions. First, television is

immeasurably more informative than newspaper and magazine stories ever were: the close-up, the instant replay, the right camera angle, slow motion. Second, television (and even radio) carries with it instantaneous commentary on what is happening, complete with statistical and personal sidelights.

Form 203

American newspapers created, nurtured and shaped mass spectator sports such as football and baseball between the end of the Civil War and World War II. Since World War II, radio and television have become more important in spreading interest and determining the form of these activities. In no other culture have the commonplace aspects of popular games so permeated daily life, even for those who pay no direct attention. Sports terminology and thinking habits have become an inseparable, often unrecognized, part of our language and even our philosophy, as in, for example, taking a rain check or getting to first base. In man-hours devoted to sports activity, in dollars generated by direct and allied sports businesses, in sheer amount of recreational attention devoted to playing, watching, reading about and listening to organized sports events, American sports have a hold on American civilization unmatched in any other time or place.

Spectator interest rests on three pillars: intimate knowledge (or at least the illusion of it) of the nature of the game and the identity of the participants, some degree of emotional alliance to one side or the other, and willingness to spend money for this sort of entertainment. Without the first, mental involvement is not possible; without the second, the first quickly becomes meaningless. But without both, the motivation for the crucial third element can't be maintained and without the commercial element the staging of sufficiently skilled and significant contests would not be possible.

The growth of television and radio, however, has meant an important difference to the first element, the spectator's knowledge. Television has changed this knowledge, that is, has increased it, in

two directions. On the one hand, it is immeasurably more informative than newspaper and magazine stories ever were. Furthermore, television (and even radio) carries instantaneous commentary on what is happening.

Form 204

American newspapers created, nurtured and shaped mass spectator sports such as football and baseball between the end of the Civil War and World War II. Radio and television have become more important in spreading interest and determining the form of these activities since World War II. The commonplace aspects of popular games have permeated daily life in our culture as in no other, even for those who pay no direct attention. Sports terminology and thinking habits have become an inseparable, often unrecognized, part of our language and even our philosophy, as in, for example, taking a rain check or getting to first base. American sports have a hold on American civilization unmatched in any other time or place, in man-hours devoted to sports activity, in dollars generated by direct and allied sports businesses, and in sheer amount of recreational attention devoted to playing, watching, reading about and listening to organized sports events.

Spectator interest rests on three pillars: intimate knowledge (or at least the illusion of knowledge) of the nature of the game and the identity of the participants, some degree of emotional alliance to one side or the other, and willingness to spend money for this sort of entertainment. Mental involvement is not possible without the knowledge; the knowledge becomes meaningless quickly without emotional alliance to a side. The motivation for the crucial commercial element can't be maintained without knowledge and emotional alliance, and without the commercial element, the staging of sufficiently skilled and significant contests would not be possible.

The growth of television and radio has meant an important difference to the spectator's knowledge, however. Television has

changed this knowledge, that is, has increased it, in two directions. First, television is immeasurably more informative than newspapers or magazines. Furthermore, television (and even radio) carries instantaneous commentary on what is happening.

Form 301

It is not easy to evaluate the performance of government. From a general point of view, performance correlates with responsiveness: the more the decision-making machinery responds to the demands and the interests articulated within the system, the higher the level of performance. By the same token, the more the government allows new socio-economic groups to participate in the system and make their interests heard and their demands satisfied, the greater the legitimacy and stability of the system. But since governmental decisions are made about the allocation of scarce resources or benefits priorities must be established. Some may get more and others less. Thus, some demands are likely to be fully met, others only in part, and some not at all. The greater the number of demands satisfied, the greater the rate of governmental performance. Thus, if defense is the highest demand, failure to be prepared against aggression would be an indicator of nonperformance. Prolonged unemployment (i.e. failure to meet the demand for full employment) would amount to nonperformance. Prolonged nonperformance would lower the attachment of many groups to the government and ultimately to the system, thus providing for instability that takes the form of a widespread rejection, not only of the government but of the system itself. Thus, failure to heed the interests and demands of a minority may account for its disaffection and the ultimate rejection of the political system on its part.

From an overall point of view, governmental performance relates to the manner in which over a given period of time the government meets and copes with specific social and environmental problems and

also anticipates them. It is not unlikely, however, that careful study might indicate that interests and individual pressures for a given decision may be difficult or impossible or even unwise to heed.

Form 302

Evaluating the performance of government is not easy. Performance correlates with responsiveness from a general point of view: the more the decision-making machinery responds to the demands and the interests articulated within the system, the higher the level of performance. By the same token, the more the government allows new socio-economic groups to participate in the system and make their interests heard and get their demands satisfied, the greater the legitimacy and stability of the system. Priorities must be established, since governmental decisions are made about the allocation of scarce resources or benefits. Some groups may get more and others less. Thus, some demands are likely to be fully met, other demands only in part, and some not at all. The greater the number of demands satisfied, the greater the rate of governmental performance. Thus failure to be prepared against aggression would be an indicator of nonperformance, if defense is the highest demand. Prolonged unemployment would amount to nonperformance (i.e. failure to meet the demand for full employment). Prolonged nonperformance would lower the attachment of many groups to the government and ultimately to the system, thus providing for instability that takes the form of a widespread rejection, not only of the government but of the system itself. Thus, failure to heed the interests and demands of a minority may account for the minority's disaffection and ultimate rejection of the political system on the minority's part.

From an overall point of view, governmental performance relates to the manner in which over a given period of time the government meets and copes with specific social and environmental problems and

also anticipates those problems. Careful study is likely to indicate, however, that interests and individual pressures for a given decision may be difficult or impossible or even unwise to heed.

Form 303

It is not easy to evaluate, that is, to judge and appraise the performance of government. From a general point of view, performance correlates with responsiveness: the more the decision-making machinery responds to the demands and the interests articulated within the system, the higher the level of performance. By the same token, the more the government allows new socio-economic groups to participate in the system and to make their interests heard and their demands satisfied, the greater the legitimacy and stability of the system. But since governmental decisions are made about the allocation of scarce resources or benefits, such as money and manpower, priorities must be established. Some may get more and others less. Thus, some demands are likely to be fully met, others only in part, and some not at all. The greater the number of demands satisfied, the greater the rate of governmental performance. Thus, if defense is the highest demand, failure to be prepared against aggression would be an indicator of nonperformance. Prolonged unemployment (i.e. failure to meet the demand for full employment) would amount to nonperformance. Prolonged nonperformance, for example, continued high unemployment and inflation, would lower the attachment of many groups to the government and ultimately to the system, thus providing for instability that takes the form of a widespread rejection, not only of the government but of the system itself. Thus, failure to heed the interests and demands of a minority may account for its disaffection and the ultimate rejection of the political system on its part.

From an overall point of view, governmental performance relates to the manner in which over a given period of time the government

meets and copes with specific social and environmental problems and also anticipates them. Interests and individual pressures may be difficult or unwise to heed.

Form 304

Evaluating, that is, judging and appraising, the performance of government is not easy. Performance correlates with responsiveness from a general point of view: the more the decision-making machinery responds to the demands and the interests articulated within the system, the higher the level of performance. By the same token, the more the government allows new socio-economic groups to participate in the system and make their interests heard and get their demands satisfied, the greater the legitimacy and stability of the system. Priorities must be established, since governmental decisions are made about the allocation of scarce resources or benefits, such as money and manpower. Some groups may get more and others less. Thus, some demands are likely to be fully met, other demands only in part, and some not at all. The greater the number of demands satisfied, the greater the rate of governmental performance. Thus, failure to be prepared against aggression would be an indicator of nonperformance, if defense is the highest demand. Prolonged unemployment would amount to nonperformance (i.e. failure to meet the demand for full employment). Prolonged nonperformance, for example continued high unemployment and inflation, would lower the attachment of many groups to the government and ultimately to the system, thus providing for instability that takes the form of a widespread rejection, not only of the government but of the system itself. Thus, failure to heed the interests and demands of a minority may account for the minority's disaffection and the ultimate rejection of the political system on the minority's part.

From an overall point of view, governmental performance relates to the manner in which over a given period of time the government meets and copes with specific social and environmental problems and also anticipates those problems. Interests and individual pressures may be difficult or unwise to heed.

BIBLIOGRAPHY

BIBLIOGRAPHY

Adler, Alfred. Understanding Human Nature. New York: Fawcett World Library, 1954.

Aitken, Kenneth. TESL Applications of Cloze Procedure: An Annotated Bibliography. ERIC: ED109922, March 28, 1975.

Anderson, Barry. The Psychology Experiment. 2nd ed. Belmont, California: Wadsworth Publishing Company, 1971.

Anderson, Jonathan. A Report of Research on Comprehension in Reading. ERIC: ED045323, August 8, 1970.

Anderson, Jonathan. "A Technique for Measuring Reading Comprehension and Readability." English Language Teaching, 25 (February, 1971), 178-82.

Best, John W. Research in Education. 2nd ed. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1970.

Bormuth, John. "Cloze as a Measure of Readability." Reading as an Intellectual Activity. Ed. J. Allen Figurel. New York: Scholastic Magazines, 1963. VIII, 131-34.

Bormuth, John. "Comparable Cloze and Multiple-Choice Comprehension Test Scores." Journal of Reading, 10 (February, 1967), 291-99.

Bormuth, John. "Literacy in the Classroom." Help for the Reading Teacher: New Directions in Research. Ed. William D. Page. Urbana, Illinois: National Council for Research in English, 1975, 60-90.

Bormuth, John. "Introduction." Readability in 1968. Ed. John Bormuth. Champaign, Illinois: NCRE/NCTE, 1968.

Botel, Morton, et al. "A Syntactic Complexity Formula." Assessment Problems in Reading. Ed. Walter H. MacGinitie. Newark, Delaware: International Reading Association, 1973. 77-86.

Bouchard, Donald, et al. Reading English as a Second Language and the Adult Learner. ERIC: ED098510, 1974.

Burt, Marina, and Heidi Dulay, eds. New Directions in Second Language Learning, Teaching and Bilingual Education. Washington: TESOL, 1975.

Campbell, Donald T. and J. C. Stanley. "Experimental and Quasi-Experimental Designs for Research on Teaching." Handbook of Research on Teaching. Ed. N. L. Gage. Chicago: Rand McNally, 1963. 171-246.

Chall, Jeanne S. Readability: An Appraisal of Research and Application. Columbus, Ohio: Bureau of Educational Research, Ohio State University, 1958.

Cherry, Colin. On Human Communication. 2nd ed. Cambridge, Massachusetts: MIT Press, 1966.

Coke, Esther. "The Effects of Readability on Oral and Silent Reading Rates." Journal of Educational Psychology, 66 (June, 1974), 406-09.

Dale, Edgar and Jeanne Chall. "A Formula for Predicting Readability." Educational Research Bulletin, 27 (Jan. 21 and Feb. 18, 1948), 11-20, 28, 37-54.

Darnell, Donald K. "Clozentropy: A Procedure for Testing English Language Proficiency for Foreign Students." Speech Monographs, 37 (March, 1970), 36-46.

Darnell, Donald K. The Development of an English Language Proficiency Test of Foreign Students Using a Clozentropy Procedure. ERIC: ED024039, 1969.

Dawkins, John. Syntax and Readability. Newark, Delaware: International Reading Association, 1975.

DeVito, Joseph. The Psychology of Speech and Language: An Introduction to Psycholinguistics. New York: Random House, 1970.

Ennis, Robert H. "Operational Definitions." American Educational Research Journal, 1 (May, 1964), 183-201.

Eskey, David. "A Model Program for Teaching Advanced Reading to Students of EFL." Language Learning, 23 (December, 1973), 169-84.

Farr, Roger. Reading: What Can Be Measured? Newark, Delaware: International Reading Association, 1969.

Ferry, Clifford L. "Discourse Coherence and Readability: A Study of the Effects of Coherence Marker Density on Reading Comprehension." Unpublished doctoral dissertation, University of Oregon, 1975.

Flesch, Rudolf. "A New Readability Yardstick." Journal of Applied Psychology, 32 (June, 1948), 221-33.

Garner, Wendell. The Processing of Information and Structure. Potomac, Maryland: Lawrence Erlbaum Associates, 1974.

Garner, Wendell. Uncertainty and Structure as Psychological Concepts. New York: John Wiley and Sons, 1962.

Gellman, Estelle S. Statistics for Teachers. New York: Harper and Row, 1973.

Gilliland, John. Readability. London: University of London Press, Ltd., 1972.

Goodman, Kenneth S. "The Psycholinguistic Nature of the Reading Process." The Psycholinguistic Nature of the Reading Process. Ed. Kenneth S. Goodman. Detroit: Wayne State University Press, 1968. 13-26.

Goodman, Kenneth S. "Reading: A Psycholinguistic Guessing Game." Current Topics in Language: Introductory Readings. Ed. Nancy A. Johnson. Cambridge, Massachusetts: Winthrop Publishers, 1976. 370-383.

Goodman, Yetta. "A Psycholinguistic Description of Observed Oral Reading Phenomena in Selected Young Beginning Readers." Unpublished doctoral dissertation, Wayne State University, 1967.

Greene, Frank P. "A Modified Cloze Procedure for Assessing Adult Reading Comprehension." Unpublished doctoral dissertation, University of Michigan, 1964.

Guilford, J. P. and Benjamin Fruchter. Fundamental Statistics in Psychology and Education. 5th ed. New York: McGraw-Hill Book Company, 1973.

Hatch, Evelyn. Research on Reading a Second Language. ERIC: ED 101590, June, 1973.

Hittleman, Daniel R. "The Readability of Subject Matter Material Rewritten on the Basis of Students' Oral Reading Miscues." Unpublished doctoral dissertation, Hofstra University, 1971.

Jongsma, Eugene. The Cloze Procedure as a Teaching Technique. Newark, Delaware: International Reading Association, 1971.

Keen, Robert. "The '20-Questions' Technique: Prediction of Visual Threshold and Measurement of Redundancy." Journal of Experimental Psychology, 100 (1973), 158-61.

Klare, George. "Assessing Readability." Reading Research Quarterly, 10 (1974-75), 62-102.

Klare, George, et al. "Cloze as a Test of Readability." International Review of Applied Psychology, 21 (1972), 77-106.

Klugh, Henry. Statistics: The Essentials for Research. New York: John Wiley and Sons, 1970.

Koppett, Leonard. The New York Times Guide to Spectator Sports. New York: New York Times Publishing, 1976.

McLeod, J. and J. Anderson. "An Approach to Assessment of Reading Ability Through Information Transmission." Journal of Reading Behavior, 2 (Spring, 1970), 116-43.

McLeod, J. and J. Anderson. "Readability Assessment and Word Redundancy of Printed English." Psychological Reports, 18 (1966), 35-38.

Macridis, Roy C. and Robert E. Ward. Modern Political Systems: Europe. 2nd ed. Englewood Cliffs, New Jersey: Prentice-Hall, 1968.

Maher, Brendan. "Language and Psychopathology." Communication, Language, and Meaning. Ed. George Miller. New York: Basic Books, Inc., 1973. 256-67.

Maminta, Rosario. "An Investigation on the Language Structures in Beginning Readers Compared with the Language Structures Taught for Oral Proficiency in the Teaching of English as a Second Language in the Philippines." Unpublished doctoral dissertation, University of British Columbia, 1969.

Miller, George. Language and Communication. New York: McGraw-Hill Book Company, 1951.

Mueller, John, Karl Schuessler and Herbert Costner. Statistical Reasoning in Sociology. 2nd ed. New York: Houghton Mifflin Co., 1970.

Nie, Norman, et al. Statistical Package for the Social Sciences (SPSS): Manual. Revised ed. New York: McGraw-Hill Book Company, 1975.

Oller, John W. "Assessing Competence in ESL: Reading." TESOL Quarterly, 6 (1972), 313-323.

Oller, John W. "Cloze, Discourse and Approximations to English." New Directions in Second Language Learning, Teaching and Bilingual Education. Ed. Marina Burt and Heidi Dulay. Washington: TESOL, 1975. 345-70.

Oller, John W. "Cloze Tests of Second Language Proficiency and What They Measure." Language Learning, 23 (June, 1973), 105-118.

Oller, John W. "Discrete Point Tests versus Tests of Integrative Skills." Focus on the Learner. Ed. John W. Oller and Jack Richards. Rowley, Massachusetts: Newbury House Publishers, 1973. 190-99.

Oller, John W. "Expectancy for Successive Elements: Key Ingredient to Language Use." Foreign Language Annals, 7 (May, 1974), 443-52.

Oller, John W. "Pragmatic Language Testing." Language Sciences, 28 (December, 1973), 7-12.

Oller, John W. "Scoring Methods and Difficulty Levels for Cloze Tests of Proficiency in English as a Second Language." Modern Language Journal, 56 (1972), 151-58.

Oller, John W. and Nevin Inal. "A Cloze Test of English Prepositions." TESOL Quarterly, 5 (December, 1971), 315-26.

Pease, Kenneth. "The Index of Contingency: A Critique." Language and Speech, 15 (1972), 58-61.

Pierce, John R. Symbols, Signals and Noise. New York: Harper and Brothers, 1961.

Romatowski, Jane. "A Psycholinguistic Description of Miscues Generated by Selected Bilingual Subjects During the Oral Reading of Instructional Reading Material as Presented in Polish Readers and English Basal Readers." Unpublished doctoral dissertation, Wayne State University, 1972.

Ruddell, Robert. "Reading Comprehension and Structural Redundancy in Written Material." Reading and Inquiry. Ed. J. Allen Figurel. Newark, Delaware: International Reading Association, 1965. X, 308-11.

Ruddell, Robert. "A Study of the Cloze Comprehension Technique in Relation to Structurally Controlled Reading Material." Improvement of Reading Through Classroom Practice. Ed. J. Allen Figurel. Newark, Delaware: International Reading Association, 1964. IX, 298-303.

Schlesinger, I. M. Sentence Structure and The Reading Process. The Hague: Mouton, 1968.

Smith, E. Brooks, Kenneth Goodman, and Robert Meredith. Language and Thinking in the Elementary School. 1st ed. New York: Holt, Rinehart and Winston, 1970.

Smith, Frank. Understanding Reading. New York: Holt, Rinehart and Winston, 1971.

Smith, Laura and Margaret Lindberg. "Building Instructional Materials." Miscue Analysis: Applications to Reading Instruction. Ed. Kenneth Goodman. Urbana, Illinois: National Council of Teachers of English, 1973. 77-90.

Spolsky, Bernard. Reduced Redundancy as a Language Testing Tool. ERIC: ED031702, September, 1969.

Taylor, Wilson L. " 'Cloze Procedure': A New Tool for Measuring Readability." Journalism Quarterly, 30 (1953), 414-33.

Taylor, Wilson L. " 'Cloze' Readability Scores as Indices of Individual Differences in Comprehension and Aptitude." Journal of Applied Psychology, 41 (1957), 19-26.

Taylor, Wilson L. "Recent Developments in the Use of 'Cloze Procedure.' " Journalism Quarterly, 33 (1956), 42-48.

Weaver, Wendell and Albert J. Kingston. "A Factor Analysis of the Cloze Procedure and Other Measures of Reading and Language Ability." Journal of Communication, 13 (1962), 252-61.

Wiersma, William. Research Methods in Education. 2nd ed. Itasca, Illinois: F. E. Peacock Publishers, 1975.

Yorio, Carlos. "Some Sources of Reading Problems for Foreign Language Learners." Language Learning, 21 (June, 1971), 107-15.