This dissertation has been microfilmed exactly as received

BARBE, Richard Homer, 1933–
AN ESTIMATION AND EXAMINATION OF THE STRUCTURAL ELEMENT OF PROSE WRITING.

The Ohio State University, Ph.D., 1961
Language and Literature, general

University Microfilms, Inc., Ann Arbor, Michigan
AN ESTIMATION AND EXAMINATION OF THE STRUCTURAL 
ELEMENT OF PROSE WRITING

DISSERTATION

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

By

Richard Homer Barbe, B.S. in Ed., M.A.

*****

The Ohio State University
1961

Approved by

Edgar Dae
Adviser
Department of Education
ACKNOWLEDGEMENTS

This project was financed by grant number D-1230 from the National Institutes of Health to the Ohio State University Research Foundation. To both of these institutions, gratitude is owed.

To the many people who contributed their time, knowledge, and efforts to this project, many thanks are due. Of note are the staff of the School of Journalism of The Ohio State University; the surgery, pediatrics, and general medicine residents of the University and Children's Hospitals; and Helen Fox, the typist who fought her way through the morass of drafts of this report.

To the men who contributed not only to this study but to the education of which this project is a part, an extreme debt is owed. This group includes Edgar Dale, from whom came background and insights in communication; Egon Guba, from whom came knowledge of research; and George Kienzle, from whom came understanding of writing and writers. These men have contributed far more than they know.

All errors and poor judgments, however, are my own contribution.
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>The Evolution of the Problem</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Structure</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Structure, Writing, and Thinking</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Structure, Logic, and Thinking</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>The Problem for Study</td>
<td>32</td>
</tr>
<tr>
<td>II</td>
<td>The Methodology for the Study</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>Instrumentation and Operational Definitions</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Experimental and Statistical Design</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Paragraph and Sample Selection</td>
<td>64</td>
</tr>
<tr>
<td>III</td>
<td>Results of the Study</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Formal Logic and Structured Writing</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Structure, Readability and Comprehensibility</td>
<td>86</td>
</tr>
<tr>
<td>IV</td>
<td>Conclusions and Implications</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Conclusions and Their Limitations</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Implications of the Study</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>Directions for Future Investigation</td>
<td>119</td>
</tr>
</tbody>
</table>
## CONTENTS (continued)

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A Test of Logical Reasoning Ability</td>
<td>122</td>
</tr>
<tr>
<td>B</td>
<td>The Nine Paragraphs Written By the Sample Subjects</td>
<td>129</td>
</tr>
<tr>
<td>C</td>
<td>The Experimental Paragraphs</td>
<td>139</td>
</tr>
<tr>
<td>D</td>
<td>Test of Comprehensibility</td>
<td>152</td>
</tr>
<tr>
<td>E</td>
<td>Technical Terms Used in the Experimental Paragraphs</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>Autobiography</td>
<td>171</td>
</tr>
</tbody>
</table>
### TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72</td>
</tr>
<tr>
<td>Extraneous Variables</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>85</td>
</tr>
<tr>
<td>Correlation Between Formal Logic and Structure</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>86</td>
</tr>
<tr>
<td>Complete Data on the Nine Paragraphs</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>89</td>
</tr>
<tr>
<td>Nonparametric Analyses of Variance on Measured Variables</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>Subject Matter of the Paragraphs</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>96</td>
</tr>
<tr>
<td>Reliabilities of the Comprehension Tests</td>
<td></td>
</tr>
</tbody>
</table>
**FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The Nine-Cell Design Matrix</td>
<td>59</td>
</tr>
<tr>
<td>2</td>
<td>Vital Statistics of the Nine Paragraphs</td>
<td>71</td>
</tr>
<tr>
<td>3</td>
<td>Distribution of Scores on the Test of Formal Logic Ability</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>The Nine Cells and Test Results</td>
<td>94</td>
</tr>
<tr>
<td>5</td>
<td>The Final Comprehensibility Indices of the Paragraphs</td>
<td>97</td>
</tr>
<tr>
<td>6</td>
<td>Number of Observations Above and Below Median</td>
<td>99</td>
</tr>
</tbody>
</table>
Chapter I

THE EVOLUTION OF THE PROBLEM

One difference between a grammarian and a writer is the size of the language units with which each is primarily concerned. Grammarians seldom treat any unit larger than the sentence. Writers, although they do deal with both words and sentences, are most often concerned with multiple-sentence units.

During the history of the English language there have been three major historical approaches to English grammar. The first of these approaches was the attempt to apply the Latin grammar to the English language. During this period, language purists attempted to purge the language of the vulgar constructions that entered the language because it was used by people other than Latin scholars. The second approach was an admission that the English language was unique and could not be bound by the rules of another language. But there remained the search for a single set of rules that could be inflexibly applied to identify "correct" English. The most recent approach recognized that English could not be bound by any inflexible
set of rules. The language, because it is a language of the people, as much as a language of the scholars, is in constant change. Grammar, then, is a matter of usage and may, at best, be generalized for relatively short periods of time.

But the analysis of writing has not undergone a similar evolution. Writing was historically treated as an art form similar in many ways to painting, music, and sculpture. This same treatment of writing is still prevalent today. In elementary and secondary schools there is an obligation to try to teach basic writing skills to all students but beyond high school there is opportunity for selection. At this higher level writing can be taught as a matter of creativity, a matter of talent.

Instruction in writing, at this level, is similar to instruction in other art forms, painting for example. To learn to write the student writes. His work is then criticized by a master, and the student writes again. This process continues until the student is so adept at creating, at expressing himself, that the master no longer finds important matters to criticize. This form of instruction is a close parallel of instruction in painting, sculpture and music.
Good writing is judged, then, on an artistic basis. There can be no numerical analysis; creativity defies quantification. Writing, as art, is judged not by rules, but by the extent to which the creator can go beyond previous work. Judgment is based on artistry and not on rigid, universal standards.

Writing, as an art, has a very important place in the culture. But there is, as well, a great need for another kind of writing. This is the kind of writing that is not meant to be artistic, but to be exact, factual and widely understandable.

A painter who creates for art lovers is called an artist, and a painter who creates for wide understanding is called a commercial artist. What is needed in writing, similarly, is the commercial writer. Perhaps, because of the nature of the material about which such a person would write, he might be better called a technical writer. The modern expansion of the amount of information being gathered in all professional fields has created a great and urgent need for this kind of writer.

There are two possible approaches to the solution of this problem. One would be to train vast numbers of people to be commercial writers. These would be people who
were adequate technical writers who were capable of dealing with professional subjects. But such a solution would create personnel shortages in other fields. Increasing each profession by the number of writers necessary would mean taking these people away from already understaffed fields.

Perhaps a better solution would be to train the present and future members of each profession to be adequate writers of their own technical material. This does not entail finding more people for each profession, but, in effect, increases the responsibility of each member of the profession.

Under this system many, but not necessarily all, dentists would be both dental scientists and adequate writers of dental papers. Not all members of the profession would have this added responsibility because some would not have information that needs to be communicated. A dental researcher, for instance, working on the frontiers of the science, would be more likely to have knowledge to be shared than the clinical practitioner.

In training these professional people to be adequate technical writers, time becomes a paramount concern. The amount of technical knowledge to be gained is constantly
increasing in each profession. Consequently, any addition to the curriculum of these professional people must satisfy not only the criterion of necessity but the criterion of economy as well. Instruction in writing must be designed, therefore, to reach the most people in the least amount of time.

To streamline the instruction of any subject, it is necessary to analyze the subject to find the few key generalizations that are most widely useful and most readily teachable. This, in part, remains to be done for the field of professional writing.

In the secondary school, where writing is taught not for artist-writers but for everyone, some of these key skills have been isolated. Grammar, spelling, punctuation and manuscript preparation are a few of these. These are stressed and drilled. The success of this instruction seems, today, to be in some doubt, but this is a side issue that will not be discussed here. These skills that have been isolated have become known as the mechanical aspects of writing. This is to separate them from the so-called stylistic elements of writing. Apparently the distinction
is that anyone can learn the mechanical aspects of writing, but only people with a certain talent can learn to achieve the quality called style.

Perhaps, too, the mechanical aspects of writing are so named because they are the elements that have been subjected to generalization and have been, to some extent, quantified. The stylistic elements, not having been so analyzed, remain in the province of the artist-writer and are nonmechanical.

The fact that these elements are still nonmechanical does not mean that they cannot be subjected to a similar analysis. It only means that they have not as yet been so treated.

Some inroads were made into these style elements in recent years by means of readability analyses. In readability analysis some of the aspects of style were generalized and quantified. These analyses measured certain readily quantifiable elements of writing, and predicted the vocabulary and idea burdens which the writing presented to the reader.

The resistance that readability analysis has met in some writing circles would indicate an unwillingness on the part of the artist-writer to admit that any elements of
style can be quantified. But this unwillingness is presented in the face of a growing body of evidence that supports the validity of readability analysis.

Two pieces of information make readability analysis easier for the artist-writer to accept. One is that the writers who have been historically treated as masters are also shown to be highly readable. Pomposity and verbosity, the faults picked out by readability analysis, are not common faults of the great writers.

Secondly, readability analysis finds its greatest usefulness and widest acceptance in commercial writing. That is, when a writer produces not for lovers of writing but for communication with a great number of people, readability analysis will be most often employed. When a writer is trying to produce a work of art he has little concern for readability analysis and, in fact, the proponents of readability analysis would not, commonly, expect him to be so concerned. But when a writer is trying to communicate specific information to a specific audience, readability scholars do expect the writer to use these principles. And, indeed, it is in this kind of writing that readability analysis is used very widely today.
No attempt will be made to discover, here, all elements of writing on which no form of mathematically related analysis has yet been tried. The following discussion and study will be limited to but one. For want of a better name, this one element will be called structure.

Structure

Defined very simply, structure is composed of all aspects of writing which deal with the order of presentation of information. It includes all elements of writing that tie sentences, paragraphs, and larger units together. The term structure is used, here, to represent those aspects of writing that teachers of writing have, variously, called unity, coherence, cohesiveness, flow, organization, order, and logic.

Most textbooks on writing include some discussion of structure under one of the names just mentioned. Some of these discussions are quite elaborate while others are more casual. But all these discussions, regardless of their elaborateness, have a common fault: they give no method for estimating the degree to which structure is
present or absent. All of these discussions agree that structure is important, many of them discuss some ways of achieving structure in writing, but none of them shows how it may be measured.

A typical discussion is this one taken from Crouch and Zetler.

Coherence, although a paragraph essential, is not so much a grammatical function as it is the smooth flow of thought within the paragraph and within the entire body of written material.... It is necessary, however, that the writer be able to tell definitely if his work has coherence. He should be able to test it to see if one thought follows another in logical order.

Apply a test for coherence to the following paragraph:...

The authors have asserted the importance of structure, they called it coherence, and they have stated that a writer should be able to test for its presence. They have also instructed the reader to apply a test for coherence to a sample. But nowhere have they given any clue as to what this test is, or, in fact, what this test might be.

\[1\] Crouch, W. George and Zetler, Robert L. A Guide to Technical Writing. p. 266.
Moore avoids the issue of testing and concentrates his discussion on the techniques by which structure is achieved. He states:

Given a well-organized paragraph, one in which the sentence ideas logically follow one after another, coherence within the paragraph is usually secured by the use of one or more of five devices: connective words, transitional phrases, repetition of key terms, pronouns looking back to antecedent nouns, and repetition of sentence patterns.

Here, coherence is defined as the tying together of sentences while the logical ordering of ideas that Crouch and Zetler called coherence is called organization. Both of these elements are parts of structure, regardless of the terms applied.

In this discussion Moore gives a clue as to what a test of one part of structure might be. If, as he says, sentences are tied together by using these five devices, then the presence or absence of these devices should show whether or not the author has tied his sentences together. But by the distinction which Moore drew, this test could apply only to the tying together of sentences and would not apply to the organization of ideas.

---

2Moore, Robert H. *Effective Writing*. p. 118.
Speaking of the aspect of structure which he called organization, Moore said:

Basic to coherence in any paragraph or in any paper, it should be understood at the outset, is clear, complete, and logical organization and development of the material. If there is no logical progression of thought from sentence to sentence, paragraph to paragraph, and section to section, coherence cannot be secured by any devices known to writing man. That should be obvious; there is no possible way to clarify non-existent relationships.3

Moore has a quite elaborate system for classifying the logical organizations that must precede coherence. He first splits organization into two types of order: natural and logical. The distinction between these is, "The natural orders...are inherent in the material itself; the logical orders are imposed by the mind of the writer on material which...might occur in any chance pattern or in none at all."4

He divides natural order into two classes, time and space. Time order is what he calls chronology and he uses the term space order for things like travel which describe movement from here, to here, to there.5

3 Ibid., p. 117. 4 Ibid., p. 78. 5 Ibid., pp. 78-79.
Moore divides the logical orders into eleven classes: climax, general to specific, specific to general, cause to effect, effect to cause, familiarity (known to unknown), complexity (simple to complex), utility (order in which reader needs to know), acceptability (pleasant before objectionable), dominant impression (impression followed by supporting details), and psychological effect (order in which reader would see ideas).^6

This discussion of the order aspect of structure also suggests a test. If these classes of orders are all-inclusive, a test might be designed to determine whether the author used one of these orders. If the author did not use one of these orders, the writing could be judged to be poorly structured.

But this test would assume that all orders are equally acceptable to readers. It would assume that any imposed (logical) order could be applied to any material. These assumptions plus the subjective nature of such a test would indicate that such a test would not be very useful.

^6Ibid., pp. 80-83.
Lorge, just before his death, began seeking a measure for structure in writing. In a mimeographed release, he showed the need for such a measure by suggesting that for all readability formulas, including his own, the readability level of writing "is invariant for any rearrangement of the sentences...".\(^7\)

He hypothesized a measure for structure by reasoning that, "If unity means that sentences follow each other in an orderly fashion, then, if any passage were to be fragmented into portions and put into random order, educated subjects should be able to reconstitute the portions into a useful order."\(^8\) Lorge used this rearrangement method on several published passages and used two different panels of judges to do the reordering. The results of measuring the amounts of agreement among judges and the agreement between the judges' consensus and the published order were compared


\(^8\) Ibid., p. 6.
From these investigations, Lorge concluded:

On the basis of these several explorations, it seems that structure can be evaluated independently of "general goodness" or "evaluative" estimations. The arrangement test which is so useful for establishing the structure in a set of pictures or cartoons seems to have utility in getting at the structure in prose. The departures from published sequence, or the departures from an order accepted by consensus may serve as the basis of a new approach for judging structure—although I doubt that it is an approach that will be relished by either teachers of composition or by expert readers.

This rearrangement test used by Lorge will be taken, in this study, as the operational definition for structure.

Judges, given the sentences of a paragraph in a random order, should find various clues to aid them in their attempt to reorder the sentences into a useful paragraph. These clues may either be inherent in the nature of the subject or be imposed onto the material by the author's use of transitional devices. Whatever form they may take, these clues represent the author's attempt to achieve structure. The extent to which a group of judges can reach agreement on an order for the sentences will measure the success the author had in this attempt. Good structure,

\[^9\text{Ibid.} , \ p. \ 7.\]
then, will be shown by a high degree of agreement among a panel of judges; poor structure will be shown by a low degree of agreement.

Defining a concept operationally is a necessary part of research, but such a definition usually does not satisfy scholars. There needs to be a more thorough description of the concept. More than how it is to be measured, there is a need to describe what it appears to be and what it appears to do.

In the previous section, structure was defined operationally as the aspects of writing that are measured by the rearrangement test. The following discussion will be an attempt to describe those aspects of writing.

As was mentioned previously, structure is a term used to replace a variety of other names for these aspects of writing. A partial list of the other names given to structure would include: unity, coherence, cohesiveness, organization and sequence. These terms have been defined variously by their users.

Coherence, for Crouch and Zetler, is "the smooth flow of thought within the paragraph and within the entire body of written material. It is an effective verbal reflection of the complete connection of one thought with the
following thought." They bring out another idea of coherence during their discussion of transition devices. They state, "... as the writer progresses from point to point in his explanation, he must notify the reader whenever he leaves one point for the next."

For these authors, then, structure is the connection between thoughts. Structure is achieved through certain writing devices such as transitional words and is a necessary part of writing from the standpoint of the reader. The elements that achieve this connection are not, by themselves, structure but are only the ways by which structure is accomplished. Structure is the connection not the connector.

Moore divides structure into coherence and organization. Coherence, for him, is accomplished by

... making clear and smooth transitions between ideas, ... making the relations between the sentences of a paragraph clear, and ... making the relations between paragraphs and even between multi-paragraph sections clear and easy for the reader to follow. The coherent paper is usually the fluent, the flowing, paper.12

10 Crouch and Zetler, op. cit., p. 266.
11 Ibid., p. 267.
12 Moore, op. cit., p. 117.
Organization is achieved by using one or more of the natural or logical orders. The sequence in which pieces of information or ideas are presented to the reader makes up the organization of the material. Order is the name Moore chose for the various classes of sequence that he was able to identify.

From the point of view of the reader, organization is a part of structure. As the reader follows the written development of the thesis, he expects certain material to follow certain orders.

The time sequence is, perhaps, the order most commonly sought by the reader in all material that involves events. The research department of the Scripps-Howard chain has some evidence to support this. They found that when news stories were rewritten into chronological or near-chronological order the readers' comprehension and recall were improved.\(^{13}\)

All of the discussion of structure thus far has been from the standpoint of ease of reading. Structure is the connection of sentence to sentence, the connection of

thought to thought, the turn and shift signals to the reader, and a sequential development that the reader understands. All of these things that structure is are designed to ease the task of reading.

One idea common to all discussions of structure bears repetition. Structure is not the elements, devices, and patterns by which connections, signals, and orders are accomplished. Structure is the result of these accomplishments.

For the writer structure is a problem, a task to be accomplished. Any measure of structure will reflect the author's ability in a particular writing skill. And as writing is more a mental process than a pattern of muscular behavior, a discussion of structure from the writer's standpoint would seem to be necessary.

Structure, Writing, and Thinking

The term writing has several meanings. A child usually begins to learn at about age five or six how to use a pencil, crayon or pen to form characters. He develops muscular proficiency in forming these characters and his report card will show a judgment of this proficiency with a mark for "writing."
After these muscular actions have become habituated, the child is called upon to use these characters, which he can now form, to represent the products of his thought. Again his report card will show a score in "writing."

But these two skills are rather different. The formation of characters is essentially a muscular activity while the symbolization of thought is basically a mental process. Neither of these classifications is completely set apart from the other, and there is a large overlapping area. But in terms of general emphasis, these meanings for writing may be treated as two separate entities. It is only the mental process definition of writing that will be used in the present discussion.

Writing coupled with its opposite number, reading, forms one of the many channels of human communication. This statement, although obvious, has significance. For if writing is a part of communication, then whatever can be said about communication generally must apply to writing.

There are many useful definitions of communication. One of these is Dewey's. He notes that the root of the term is the same as the roots of common, communion and community. This indicates that communication must have qualities of sameness, sharing, and agreement. Communication, then,
becomes a process of sharing thoughts and ideas toward the goal of reaching a mutual understanding or meaning; "... consensus demands communication." Writing must be a means through which this process may be accomplished.

Communication is cyclic. All parts of and parties to a communication system are gathered up into a transactional whole. For purposes of examination, though, this whole may be divided into parts which are more easily described.

The message to be communicated is placed into its transmission channel by the communicator. The formation of the message and its placement into the writing-reading channel would both seem to be parts of the writing process. That is, the writer both conceives the message and symbolizes it.

Two other communication-related concepts need definition before structure can be identified as a part of this writing process. The receiver is the person who stands at the other end of the channel. In the case of the writing-reading channel, the receiver is the reader. Noise is interference that stands between the communicator and the

---

receiver. It prevents the message as obtained by the receiver from being identical to the message as transmitted by the communicator.

Structure is an aspect of the written message. As defined in the previous section from the reader's standpoint, structure reduces the noise between the author and the reader.

From the viewpoint of the writer, the communicator, structure is a quality of the message that he must achieve. It is an effort, on his part, to reduce the noise level for the reader.

The preparation of the message is a mental activity. Placing the message into the written channel is also a mental activity. If the term is used loosely, such mental activity can be broadly called thinking. This follows from such definitions of thinking as Guilford's. He indicated that "thinking is the broad term that includes all symbolic behavior."\(^{15}\)

While there is very little discussion in the literature of the relation between writing and thinking, there is

an abundance of discussion about the relationship of thinking to writing's counterpart, reading. One of the earliest of these that is still often cited is Thorndike's article published in 1917. Three of the most recent of these are Buswell's study published in 1951, and Betts' studies published in 1956.

A few people have approached the problem of the relation of writing to thinking. Sams decided that "writing may be envisioned as a mode of thought," and that "Paper and pencil will teach him what and how to think." 


17 Buswell, Guy T. "Relationship between Rate of Thinking and Rate of Reading," School Review, LIX (September, 1951), pp. 339-346.


Michaels assumed that thinking was a necessary adjunct to writing, and he chose this relation as a focus for his criticism of elementary and secondary school training. He wrote, "The unfortunate fact is that, while college freshmen are often unprepared in the simple elements of English, they are even more abominably incapable of critical thinking."  

There seems to be general agreement that there is a relationship between writing and thinking because both are symbolic behaviors. If this statement shows how the two are related, it might be fruitful to point out their differences.

The first difference, obviously, is the scope of each. Since thinking encompasses all symbolic behavior, it necessarily includes such behaviors as reading, speaking, listening, viewing and picturing. Writing, while related to these through its relation to thinking, is uniquely different from each of these. It is a process occurring before transmission of the message. This makes it different from reading, listening and viewing which occur at the reception end of the channel. Writing uses

\[^{21}\text{Michaels, H.S. "Logical Approach to Composition," College English, XII (April, 1951), p. 390.}^{21}\]
language symbols; it differs from picturing which uses signs or nonlanguage symbols. And writing requires a visual channel. This distinguishes it from speaking. Writing may be a part of thinking, but the two are not synonymous because thinking includes symbolic behaviors other than writing.

Thinking is, also, more than the sum of these communicative behaviors. Communication is usually limited, by definition, to situations involving more than one human being. Man-animal or man-object communication situations and man's communicating with himself are normally disallowed. These possibilities are symbolic behaviors which are noncommunicative. Thinking, then, includes more than symbolic, communicative behaviors; it also includes non-communicative behaviors like introspection. Again, thinking includes behaviors other than writing.
Writing has a tangible result. This may be both a difference and a similarity between writing and thinking.

As Dewey wrote:

The mere occurrence of ideas or suggestions constitutes thinking, but not reflective thinking.... Only when the succession is so controlled that it is an orderly sequence leading up to a conclusion that contains the intellectual force of the preceding ideas, do we have reflective thought. 22

These two kinds of thinking, then, give a basis for both a similarity and dissimilarity between writing and thinking. Writing has a product. Thinking which produces no conclusion, nonreflective thinking, differs from writing on this point. Reflective thinking is, on the other hand, similar to writing in that both involve the production of a result.

To summarize, we can conclude, then, that writing is similar to thinking in that both are symbolic behaviors of man. Thinking has a broader scope than writing because thinking includes communicative behaviors other than writing and it includes noncommunicative symbolic behaviors as well. Thinking differs from writing, also, with regard to their products. There is a particular kind of thinking,
reflective thinking, which is similar to writing on this point; it has a product. Nonreflective thinking has no product.

Writing cannot be thought to be completely different from thinking nor are they completely alike. The two appear to be related but not synonymous.

Writing's communication function is not directly a part of thinking. Thinking is personal as opposed to communicative. But writing may affect thought. As Dewey said in a discussion of the teaching of language:

That problem is to direct pupil's oral and written speech, used primarily for practical and social ends, so that gradually it shall become a conscious tool of conveying knowledge and assisting thought.²³

Conversely, thinking produces the material for writing, and, to this extent at least, thinking affects writing. Perhaps the most acceptable conclusion, then, is that writing and thinking are intimately related in an interacting sense. Each maintains certain unique features, but each is related to the other.

²³Ibid., p. 239.
In this section, discussion of the patterns or the processes of these two concepts, thinking and writing, has been avoided. These processes will be treated in the following section.

**Structure, Logic and Thinking**

In his statement defining the difference between thinking and reflective thinking, Dewey used such terms as "succession" and "orderly sequence." This would seem to mean that reflective thinking follows a pattern. In earlier sections such a pattern for writing was called structure. Perhaps, then, these two processes follow similar paths of operation.

From the preceding paragraph, there comes sufficient evidence to support a belief that both thinking and writing are patterned and that the patterns for both seem to

---

evolve from the life experience of the person involved.

About the structure of writing, Crouch and Zetler wrote:

Most of us have this technique to some degree, although we have usually learned it from experience rather than through formal teaching. We have observed logical sequences in actual life; we automatically apply them to writing if our observation has been exact.25

Dewey discussed the development of order in thinking by saying that men used their occupations as the proving ground for thought patterns. He said, "... all people at the outset, and the majority of people probably all their lives, attain to some ordering of thought through ordering of action."26 Both structure of writing and the pattern of thinking seem to share a similar development.

The term logic appears in most discussions of patterns. It could hardly be otherwise for logic is, simply, the generic term for the system or pattern of a process. This is the denotative meaning for logic. But logic has a connotative meaning of syllogistic reasoning. In this discussion these meanings will be differentiated by using logic for the denotative meaning and formal logic for the connotative meaning.

25 Crouch and Zetler, op. cit., p. 266.

26 Dewey, op. cit., p. 49
Dewey discussed, at some length, the relation between formal logic and thinking. Formal logic operates in thinking much as a map does for an explorer. It is a product, like a map, of previous inquiry and serves future thought, like a man's future explorations, as "a check on his position and a guide to his movements. But it does not tell him where to go; his own desires and plans determine his goal, as his own past determines where he is now and where he must start from."  

Dewey pointed out specific differences between formal logic and thinking.

Inspection shows important differences between formal reasoning and thinking as it actually goes on in the mind of a person. (1) The subject matter of logic is strictly impersonal, as much so as the formulae of algebra.... (2) The forms of logic are constant, unchanging, indifferent to the subject matter with which they are filled.... (3) Because forms are uniform and hospitable to any subject matter whatever, they pay no attention to context.  

Thinking, then, is logical but not formally logical.

And the logical aspects of thinking have importance in the teaching of writing. In discussing the process by

---

27 Ibid., p. 74. 28 Ibid., p. 72.
which education must transform language into an intellec-
tual tool, Dewey said:

The difficulty lies in making over habits that have to do with "ordinary affairs and conven-
iences" into habits concerned with "precise notions." The successful accomplishing of the transformation requires (a) enlarging the pupil's vocabulary, (b) rendering its terms more precise and accurate, and (c) forming habits of consecutive discourse. 29

The "habits of consecutive discourse" are the means by which, in writing, structure is achieved.

No relation, other than similar development and similar reasons for existence, has been shown between logic in thinking and structure in writing. But a more direct relation can be shown for formal logic and writing.

As was mentioned earlier, Dewey indicated that a syllogistic system was a product of thought that was useful for future thought. In this respect, he believed the formal logic to be useful, also, in communication. He said:

In short, these forms apply not to reaching conclusions, not to arriving at beliefs and knowledge, but to the most effective way in which to set forth what has already been concluded, so as to convince others (or one-
self if one wishes to recall to mind its grounds) of the soundness of the result. 30

29 Ibid., p. 240. 30 Ibid., p. 74.
Here, then, is a direct relation between an aspect of thinking and the structural aspect of writing. A set of syllogisms and conclusions, a formal logic, is a product of the thinking process. This product is not only useful for future thinking, but it is the most effective structuring of the message used to communicate the product to others. To this extent, at least, structure in writing is a direct reflection of the thinking which preceded the writing.

But this relation has been, at best, postulated. It remains to be empirically demonstrated.

In the preceding sections of this chapter a number of other relations and concepts were postulated. These, too, require demonstration. The following section will describe these relationships more specifically in discussing the problems for study in this investigation.
The Problem for Study

Teachers of writing are faced with the problem of teaching two quite different writing skills. They must teach the mechanical aspects of writing such as spelling, grammar, punctuation, and manuscript preparation. They must also teach the nonmechanical aspects of writing. These are the style elements of writing which are usually considered unmeasurable.

Some style elements, however, have been quantified, and still another element appears to be capable of quantification. Readability, for example, measures several of these elements in order to predict the difficulty that the material presents to a reader. And now enough evidence is available to support ideas about the measurement of the structural element.

The Lorge rearrangement test of structure is not sufficiently well developed at present to permit its use in definitive research. This fact limits the research proposed for this investigation to a feasibility-type study. This does not mean that specific hypotheses will not be formulated and tested. Nor does it mean that the investigation will not be valuable. What it does mean is that the
first, and most basic, question that will be answered will be about the feasibility of using this rearrangement test as a measure of the structural element of writing.

The point has already been made that the structural aspect of writing is important as it affects the ease of understanding the material. That is, it affects the comprehensibility with which the material will be read. This has been postulated, but not demonstrated. The experimental demonstration of this effect will be the first of the specific problems for this investigation.

Readability is the only other style-related concept that has been measured so far. Lorge concluded on the basis of his study that structure could be isolated and measured independently from the "general goodness" of the writing. But can structure be isolated from the elements measured by readability? The obvious answer is that it can because the two concepts are operationally so different. A readability score will remain constant for all orders of sentences, and the order of presentation of ideas may remain the same at all levels of readability. But in the effect of these style elements on the comprehensibility of a piece of writing the two may interact. An experimental
examination of the relation between these two elements will be the second specific problem for study.

The last specific problem for investigation is the relation between thinking and structure. As was mentioned earlier in this chapter, Dewey believed a formal logic system of premises and conclusions to be both a product of thought and an effective structure for communication. Tests of a person's ability to develop a formal logic pattern have been constructed by other researchers and some validity has been shown for these tests. If the rearrangement test is a feasible measure of structure, it must also measure the structure Dewey believed most effective, namely, formal logic. Structure is both an element in writing and a product of the writer. Perhaps, then, the ability of the writer to form a premise-conclusion pattern will relate to his ability to structure his writing. This possible relationship is the last specific problem for study.

The evidence presented in previous sections of this chapter seems sufficient to warrant the formation of specific hypotheses for these problems. This is possible for the three specific problems but not for the question of feasibility. This question will carry through the
whole of this study as a normative question seeking an indicative, as opposed to quantitative, answer.

**Hypothesis 1:** The abilities of people to establish premise-conclusion logic patterns will show a relation to their abilities to structure their prose writing that is statistically significant beyond the five per cent level. This hypothesis follows from the third specific problem listed for this study. The similarity of these two tasks will support a belief that if people differ in their abilities to perform these two tasks, the two abilities will be directly related. It should be pointed out that the test of this hypothesis will show relation and not causation.

**Hypothesis 2:** Structure, as measured by the Lorge rearrangement test, will have an effect on the comprehensibility of writing that will be statistically significant beyond the five per cent level. This hypothesis follows from the evidence and intuition that the order of presentation will affect the reader's ability to grasp the information which the material contains.

**Hypothesis 3:** Readability will have an effect on the comprehensibility of writing that will be statistically significant beyond the five per cent level. This hypothesis
is added here for two reasons even though it has been well supported by other research. First, any design that permits testing the other hypotheses of this study will also permit testing this hypothesis. This will add one more bit of evidence to the already large amount relating to this hypothesis. Second, the testing of this hypothesis will provide an internal check on the design for this investigation. If this hypothesis is not supported by the data gathered, some methodological error will be indicated. This effect of readability on comprehensibility has been demonstrated often enough that contradictory results, here, would point out an experimental error.

**Hypothesis 4**: Structure and readability will show no statistically significant interaction effect. This hypothesis indicates a belief that these two concepts act independently in their effect upon the comprehensibility of writing.

In order to test these hypotheses an experimental design and methodology must be developed adequate to this task. Also, reliable and valid instruments must be either found or developed which will provide adequate operational definitions of the concepts involved.
Assuming that these hypotheses can be adequately tested, what does this study mean for the very practical functions of writing and instruction in writing? What kinds of information will this study provide that are not presently available?

The question overriding this entire study is the feasibility of this method of analyzing and measuring the structure of writing. Regardless of the outcome, this study will provide an indicative answer to this question. If this method does prove to be feasible, then the implications for writing and writing instruction are very important. A basic problem facing professions today is not the training of a few great writers but, rather, the training of many adequate writers. An important step toward the solution of this problem is a thorough analysis of the material to be taught. This method of analysis, if proven feasible, will provide a basis for doing this.

Again assuming feasibility, pieces of writing could be found or prepared using this method of analyzing structure so that they vary in their structural aspects. These could then be given a more thorough analysis to find out why they vary. In earlier sections of this chapter where
structure was defined and described, very little attention was given to its causes. Questions still remain as to how structure is accomplished, how it may be improved, what effect changes in structure would have on comprehensibility, and what verbal or grammatical elements are involved. The answers have been hinted at but have not been completely defined. Such definitive answers require a thorough examination of materials whose structures are known to vary.

This method of measuring structure, even if feasible, is unwieldy. But with materials of known structure available, it would be possible to develop easier, more objective methods of measurement. It is possible that a formula similar to a readability formula might be one such method. In order to achieve such a measure, materials must be available against which the measure may be built. The method used in this study, if feasible, may provide a means for developing such materials.

Before all teachers of writing are completely alienated by these possibilities, it should be stated, again, that the concern here is with adequately communicable writing, not artistic, beautiful or great writing.
Our concern is with helping the professional man share his knowledge, and not with trying to reduce all prose to a series of mathematical equations.

The specific hypotheses to be examined in this study have implications for writing and for instruction in writing. The relation between formal logic and structure in writing is especially important. It may be, for example, that members of a profession such as dentistry who enter fields where they are called upon to write, like research or teaching, are those who have these abilities. That is, these abilities may be elements of the natural selection process by which men become researchers or educators. It seems that one ability which enters into the selection of a research career, for example, might be this formal logic ability. Thus, researchers may be, already, well able to structure their writing.

If a relationship can be shown between these two abilities, a causal relationship might be established. It may be that training in research methods will improve the structure of the student's writing. Or the reverse, training in production of well-structured writing may improve the student's ability as a researcher.
The hypothesis about the relation between structure and comprehensibility is also important. If this relation exists and structure is shown to be separate from readability, then a whole new element of writing will have been isolated. This would have the effect of taking one more aspect of writing out of the conglomerate category called style. This would mean that structure would be an identifiable entity which would be susceptible of discussion and study.

This study has an important part in the examination of these possibilities. These are, perhaps, the most important possibilities, but there may be many others. Specific implications of this study will be more fully discussed later in connection with the experimental results.

The obvious first step in such a study is to develop an experimental design and methodology adequate to the task. This development will be the topic of the next chapter.
Chapter II

THE METHODOLOGY FOR THE STUDY

This chapter will discuss the development of the instruments and methods used to conduct this investigation. The topics to be discussed will include instrumentation, sample selection, and experimental and statistical designs. The limitations placed on the study by the decisions made under each of these topics will also be given attention.

An experiment is always a compromise between an ideal and the practical. Perhaps one measure of the quality of research is the extent to which the ideal is achieved within the limits of the practical realities of the situation. This study, like all research, is limited by such practical considerations as availability of instruments, suitable subjects, and time. These must be considered when any decision is made concerning the plan by which the study is to be done.
There are four concepts which are basic to this study: structure, readability, comprehensibility, and formal logic ability. All require measurement. The first section of this chapter, then, will discuss the location or preparation of measures which will provide suitable operational definitions of these concepts.

Instrumentation and Operational Definitions

Structure is the concept about which this whole study is centered, and it deserves to be treated first.

The operational definition of this concept was partially discussed in the first chapter. Lorge developed a method by which he believed the structure of writing could be measured. His method was to arrange the fragments of a piece of written material in a random order and give them to a panel of judges. The judges were asked to rearrange these fragments into a sensible order. The measure of structure was the extent to which the judges achieved the original order or an order agreed upon by consensus, or

1Lorge, op. cit.
structure was the extent to which the judges agreed among themselves on what the order should be.

The measure of the extent to which the judges achieved an accepted standard sequence was a simple observational measure based on the number of units that were displaced in order. The mean of the numbers of these displacements made by all judges was the measure of the extent to which the judges achieved the standard order, the measure of the writing's structure.

This system has a certain logical validity. If structure is thought of as the bonds between fragments of writing, then each unit must have a clear bond connecting it with the fragment preceding it in the sequence. The maximum number of these bonds necessary to achieve good structure is one less than the number of fragments because the first fragment has no predecessor.

The number of displacements of units that a judge makes, then, is a count of the number of these bonds that were either nonexistent or that were not apparent to him. Good structure, under this system, is measured by the number of bonds that existed and were apparent to the judges. Good structure would be reflected in a minimal mean number of displacements.
This system has two major limitations. It assumes that a valid standard order can be selected. At the present state of knowledge about structure, this assumption is extremely doubtful. In most writing, there would be some question as to the correctness of the author's sequence. Disregarding the notion of correctness, it is doubtful whether we may ever assume that the author's sequence was the best order.

The judge's consensus can be more easily discarded as a standard. Sequences, which when taken individually may appear reasonable, may make no sense at all when taken as a collective average. If half of the judges placed a fragment second and half placed it fourth, the average position would be third, a position no judge believed the fragment should hold.

Perhaps Lorge meant by consensus, not an average but a determination of the majority opinion expressed by the judges. This would eliminate the objection to averaging positions, but it would add an objection. Using the majority opinion as the standard sequence does not allow for consideration of the minority opinion. Judges in the minority must just be disregarded.
A second major limitation on this measurement system is that it assumes that the lead fragment is easily identified. The lead fragment has no predecessor and has bonds only to subsequent material. This fact is not adequately accounted for by this measurement system which counts only bonds to preceding material.

The one system of measurement suggested by Lorge that seems to overcome these difficulties is the determination of the degree to which the judges agree among themselves. The basis for this system is the assumption that good structure will result in a high degree of agreement.

There is a statistic available that measures this agreement. Kendall's coefficient of concordance ($W$)\(^2\) is much like a multiple correlation coefficient. It shows the extent to which ordinal data are placed into the same rank-order by two or more raters.

Using this statistic to measure structure assumes that if writing has good structure judges will see it and will tend to agree upon a sequence. Material, then, would have poor or no structure when the fragments could

\[^2\text{ Siegel, Sidney, Nonparametric Statistics for the Behavioral Sciences, pp. 229-238.}\]
be equally well ordered in any sequence. This poor structure would be reflected in a low agreement, a low coefficient of concordance.

This system makes no assumptions about the prior existence of a "best" order; it makes no assumption about an average sequence or a consensus order. And it includes the agreement of the judges on the lead fragment in its measurement. This system, therefore, seems to be the most acceptable for this study, and it will be used, here, as the measure and operational definition of structure.

Structure, then, is the quality of writing measured by the rearrangement test used with the Kendall coefficient of concordance.

Readability and comprehensibility are two concepts that have been extensively discussed and well defined operationally by others. They need to be discussed here only to point out the specific meanings that each has in relation to this study and to show how the two concepts differ in meaning.

Readability, as it is defined by Dale and Chall, is the sum of all elements in a piece of writing that

---

affect the success that readers have in understanding what they read. This definition would seem to include the elements of structure which, also, have been suggested to have an effect on the reader's understanding. But operationally there is a difference.

All readability formulas measure a few easily quantifiable elements of writing, such as average sentence length, number of personal pronouns, number of prepositional phrases, percentage of polysyllabic words, percentage of monosyllabic words, or number of words not found on a standard list. The difficulty of the material is predicted from these measurements. This prediction is of the form: this particular audience (educational level group) can be expected to average a particular comprehension score on this particular piece of writing. Invariably good structure is assumed.

The relation among these concepts, then, is: assuming good structure, readability scores predict the comprehensibility scores of a particular group of readers for a particular piece of written material.
In this study, readability will be measured and operationally defined by the Dale-Chall Readability Formula. This formula is preferred because (1) it has been shown to be highly reliable; (2) it has a high validity; and (3) it is widely known and used. For this investigation, then, readability is what the Dale-Chall formula measures.

The comprehensibility scores achieved by a particular group of readers on a particular piece of writing are predicted by readability scores. Comprehensibility, then, is that aspect of writing that affects the amount of information that readers obtain by reading. It is most often measured by designing questions covering the range of information contained in a piece of writing and asking these questions of the reader immediately after he has finished reading. There can be no standard test of comprehensibility because the questions must relate to a specific piece of writing.

---

4Dale, Edgar and Chall, Jeanne S. A Formula for Predicting Readability.
Time also seems to be a factor in comprehensibility. It seems reasonable to believe that if a reader is given sufficient time, sooner or later he will be able to obtain and understand all the information contained in the writing. Usually a reader is scored on his reading ability in two areas, speed of reading and comprehension. In dealing with the comprehensibility of the writing, then, it appears logical that both time and understanding should be taken into account. Thus, comprehensibility will be defined as the number of questions a reader answers correctly per unit of reading time. This is the comprehensibility for one reader. Taking the median of these scores over a sample group of readers will give the median comprehensibility score for the material. This is the score which will be used in this study.

Comprehensibility, then, will be measured by asking a sample group to read the material, recording their reading times, and scoring their ability to answer informational questions about what they have read. The comprehensibility score will be the median, taken across the sample, of the products of the percentages of questions answered correctly and the reading rates in words per second.
The actual writings used, and the test questions prepared for them will be found in Appendices C and D.

**Formal logic ability** has been partly discussed in the preceding chapter. This ability was, there, defined as the ability of an individual to prepare a syllogistic set of premises and conclusions from available evidence.

The Morgans designed a measure of this ability. Using this test, they found this ability to be differentially present in adults. They found this ability to be influenced by training, but to be present, to some degree, in all the adults they tested. They agreed with Miller who felt

that logic is a *formal* system, just as arithmetic is a *formal* system, and to expect untrained subjects to think logically is much the same as to expect preschool children to know the multiplication table.

5Morgan, W. J. and Morgan, A. B. *The Morgan Test of Logical Reasoning*.


The Morgans concluded that all adults receive informal training in this ability, but some, who receive formal training, have this ability to a greater degree.

For this study formal logic ability will be measured by a test developed specifically for this study. The Morgan test is not acceptable because it requires too much time to administer and it uses several symbol systems. That is, some items are designed around nonword symbols. The form of the questions in the test designed for this study, however, will be similar to that used in the Morgan test.

The subjects will be given two or more premises and one or more related conclusions. They will be asked to indicate whether each conclusion is a valid or invalid part of the logic pattern. For example, given the premises that (1) some dogs have stubby tails and (2) my dog has a stubby tail, the conclusion, my dog is some dog (!), should be marked invalid.

This ability would seem to be a function of time. That is, the person who is able to determine very quickly whether a conclusion is valid or not would seem to have this ability in a greater degree than the person who takes
longer to decide. The test of this ability will, therefore, be a timed test.

Formal logic ability will be measured and operationally defined by the percentage of test items answered correctly in a given time. The actual test used is included, here, as Appendix A.

**Experimental and Statistical Design**

This section will discuss the experimental and statistical procedures to be followed. This will be an outline of how the instruments discussed in the preceding section will be used to provide data to answer the feasibility question and to test empirically the four specific hypotheses.

The question about the feasibility of this method of measuring structure will underlie the entire design. The answer to this question will be based, in the greatest part, upon the judgment of the researcher. This judgment will be largely subjective, but there will be some data obtained that will provide some support. For example, evidence which shows that the quality measured by the rearrangement test has an effect on comprehensibility and
has no interaction with readability would tend to support the feasibility of this method. Evidence to the contrary would not support the feasibility.

But there will be no specific test or procedure that will objectively answer the question of feasibility. All aspects of design that are to be used to test the specific hypotheses will bear upon this answer, but the answer will still be largely a subjective judgment.

The four specific hypotheses to be tested require two types of experimental design. The hypothesis dealing with the relation between formal logic ability and the production of well-structured writing requires only a correlational design. The other three hypotheses require a more sophisticated design, one that will permit the demonstration of the primary and the interaction effects of structure and readability on comprehensibility.

The procedures to be followed to test the relation hypothesis will be the same as those followed for any correlational study. A sample group of subjects will be measured on both abilities and the results will be statistically correlated.
In this study, the sample will be given the test of formal logic ability. After they have finished this test, but at the same meeting, the group will be asked to write a paragraph. The topic for this paragraph will be, generally, the same for all members of the sample. That is, they will be given a specific topic on which to write, one on which all members of the sample will have a large amount of information. The topic and the instructions will point up the idea that what is desired is a logically arranged, well-developed argument.

An argumentative topic will be selected so that the members of the sample will have the greatest possible latitude in showing their ability to structure their writing. Narration and description, while they too must be structured, do not give the easy opportunity for structuring that argument provides.

An argument is, however, limited by the amount of evidence that the writer possesses. But, here, the topic selected will be general enough so that in choosing his approach each writer will have to discard evidence rather than try to find enough. The limitations of evidence, then, will not be applicable and will not significantly influence the results.
It may sometime be possible to design a multiple-choice or true-false test that could measure an individual's ability to structure writing. But this was not attempted for this study. Indeed, at the present level of knowledge about structure it hardly seems possible to design such a test. At least one English exercise book, though, contains exercises which purport to do just this. The Scott-Foresman book of exercises, Guide to Modern English, contains problems in which the student is asked to rearrange randomized sentences. Thus, his ability to structure his own writing is "measured" and, with sufficient practice, improved. The assumption is, apparently, that practice in reordering sentences written by others will improve the student's ordering of his own sentences. This certainly has some logical merit, but there can, as yet, be no experimental support of this teaching technique for there is, as yet, no adequate measure of structure.

There is a limitation placed on this procedure. The size of the sample necessary to give adequate data on the quality of the formal logic test is too large to be practical for the panel of judges to handle. The availability of subjects and the number of subjects necessary
to give reliable data about the formal logic test indicates that a sample of about fifty subjects will be sufficient. But the time which the judges have available and the other judging tasks asked of them limit the number of paragraphs which they can handle to about ten.

This limitation forces the procedure to be slightly modified. Instead of correlating all formal logic test results with the structure scores on all paragraphs, about ten paragraphs will be selected. These will be selected so that the authors' scores on the formal logic test encompass nearly the entire range of scores. These ten structure scores will be correlated with the appropriate ten formal logic scores using the Spearman rank-order correlation (rho) technique.  

Four other qualities of these paragraphs and of the sample subjects will be controlled. The number of sentences, the number of words, the amount of formal logic training, and the amount of training in writing would all influence this correlation and will all be controlled in the selection of the ten paragraphs. These paragraphs will

---

be chosen so that there is no significant pattern of these four qualities in the selected paragraphs.

The significance of the resulting correlation will be based on ten scores versus ten scores. Even though the ten paragraphs were selected so that a wide range of formal logic scores was represented, no statistical justification could be found to support treating the correlation as representing the entire sample (about fifty) versus the entire sample. This has the effect of increasing the correlation needed to show statistical significance.

These experimental and statistical procedures will test the null hypothesis: no statistically significant correlation exists between formal logic ability and the ability to produce well-structured writing. The area for rejection of the null hypothesis will be beyond the five per cent level of significance.

The remaining three hypotheses, as noted earlier, will require more sophisticated experimental and statistical designs. But these three are similar enough that one well-designed experiment will test them all.
There are, in these three hypotheses, two independent variables, readability and structure. The dependent variable to be measured against them is comprehensibility. The three relationships to be demonstrated are (1) the effect of readability on comprehensibility, (2) the effect of structure on comprehensibility, and (3) the lack of an interaction effect between readability and structure.

A nine-cell matrix design using three levels of readability and three levels of structure will provide a test of the three hypotheses. Paragraphs will be either found or prepared so that each satisfies the structural and readability criteria for placement in one of the nine matrix cells.
Figure 1 shows the design matrix.

Figure 1

The Nine-Cell Design Matrix

<table>
<thead>
<tr>
<th>Readability</th>
<th>Structure</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>Fair</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Each cell will contain one paragraph, and each paragraph will vary from the other eight in its combination of structure and readability qualities. That is, the paragraph in cell 1 will have poor readability and poor structure; the one in cell 5 will have fair readability and fair structure; and the one in cell 9 will be good in both variables. Poor and good are used only to indicate relative levels of the variables and are not meant to be judgments of these levels.

Certain extraneous variables will be either randomized throughout the design or otherwise controlled. These include number of words per paragraph, number of sentences
per paragraph, subject matter of the paragraph, and the percentage of technical terms in each paragraph.

This last item, percentage of technical terms, is included because there is no present agreement among readability scholars as to the effect of these terms on readability. These terms inflate the readability score because they are generally long words and generally not common in lay language. They may be so common in the language of the particular profession that they do not add to the reading difficulty of the material. As a result of this unanswered doubt, this variable will be controlled in this design.

The comprehensibility of these nine paragraphs is the dependent variable to be measured, and a comprehension test will be prepared for each. These tests will be fairly short and will consist of modified true-false questions.

The tests will be kept to a minimal length consistent with reliability. This will be done so that all nine paragraphs and tests may be administered to the sample subjects in one testing session thus avoiding the confounding errors of such variables as subjects' fatigue and attitude.
The item form will be modified true-false. Modified true-false means, here, that the statement is true only if it is an accurate paraphrase of a point made in the paragraph. The statement is false if it is either an inaccurate paraphrase or if the information was not contained in the paragraph. That is, a statement which is an accurate statement of fact is still to be marked false if that fact was not a part of the paragraph. This should keep the tests from being measures of prior knowledge. It should force the subjects to respond on the basis of what they read not what they knew before reading.

This item form is preferred for two other reasons. First, it is more quickly answered and scored than are unaided recall items. And, second, this item form is more easily prepared by an investigator with only a basic knowledge of the paragraphs' subject matter. To design good multiple-choice items, for example, the test designer must have sufficient knowledge to write distracting choices which are superficially reasonable but which are incorrect on the basis of the information in the paragraph. This high level of knowledge is not required in order to write adequate true-false items.
The actual paragraphs used in this design and the comprehension tests prepared for them are included as Appendices C and D.

The sample subjects will read each paragraph and will respond immediately to the appropriate comprehension test. Their reading times will be recorded. The comprehensibility of the paragraph, then, will be the median, taken across the entire sample, of the products of percentages of items correct and the reading rates in words per second.

One other extraneous variable will be randomized by the way in which the tests are administered. Each subject will read and be tested on each paragraph in a different sequence. The whole sample will not do the same paragraph first and the same one last, etc. This procedure randomizes throughout the design such order variables as fatigue and practice.

The comprehensibility scores for the nine paragraphs will be statistically analyzed to test the three related hypotheses.

The ordinal nature of the data precludes using parametric statistics. There are several nonparametric methods available for analyzing data such as this, but
nearly all assume that there is no interaction between the major variables. They assume this, but they provide no way of testing this assumption. This lack of interaction can not be assumed because one of the hypotheses being tested is that there is no significant interaction. Wilson, however, developed a statistical method that does provide a measure of interaction. This statistic will be used.

The following null hypotheses will be tested using the comprehensibility scores as the basis for statistical analysis.

1. There will be no significant variance between the rows of the matrix (readability). The area of rejection will be that beyond the five per cent level.

2. There will be no significant variance between the columns of the matrix (structure). Again, the area of rejection will be that beyond the five per cent level.

---

3. There will be a significant variance attributable to the interaction between the two major variables. The area of rejection will be that below the five per cent level.

This last hypothesis is difficult to write in null form. If it should happen that there is a significant interaction, this would mask the primary effects of readability and structure. That is, there must be no significant interaction or else the significance of the primary effects can not be estimated.

The statistically probable rejection of these null hypotheses will be the basis for supporting the relations predicted theoretically for these concepts.

**Paragraph and Sample Selection**

The nine paragraphs selected must meet a rather exacting set of criteria. The three by three design was chosen over larger matrices because these criteria make it so difficult to find or to prepare acceptable paragraphs. Using more levels of readability and structure would improve the refinement of the results, but within the limitations of time and adequacy the 3x3 design was chosen.
Each paragraph, first, must contain the levels of structure and readability for the cell in which it is to be used. This means that each of the nine paragraphs must have a combination of these qualities different from all the other paragraphs. Expressed as readability-structure, there must be a poor-poor, a fair-poor, a poor-fair, a fair-fair, etc.

In addition, there may be no significant pattern, either horizontally or vertically, of number of sentences, number of words, percentage of technical terms, or subject matter. Also, each paragraph must contain enough information to permit reliable testing of comprehension.

These criteria make the selection task difficult.

The most appropriate method of selection would seem to be to choose from published journals a series of more-or-less independent paragraphs all of which contain approximately the same number of words and sentences. Enough of these will be selected so that the chances of their meeting the criteria of structure and readability are good. These paragraphs will have to be relatively long in order to contain a sufficiently large amount of information.
Any limits placed on length or number of sentences would, necessarily, be arbitrary. The units actually chosen were eight to ten sentences and one hundred to two hundred words.

A search of the medical and dental journals of recent months produced eight paragraphs that satisfied these criteria. In addition, two other paragraphs based on summaries of journal articles were especially written.

Measurement of the readability levels of these ten paragraphs with the Dale-Chall Readability Formula showed that eight of these paragraphs fell, rather neatly, into three different readability classifications. Four had relatively good readability; two had moderately good readability; and two had relatively poor readability.

These eight paragraphs were divided into their component sentences and were given to the panel of judges for rearrangement to measure their structural levels.

At this point, good fortune took a hand. Seven of these eight paragraphs proved to have combinations of readability and structure that permitted their immediate placement into seven of the nine matrix cells. This was a fortunate event that would probably not happen again.
All that remained, then, was to fill the remaining two matrix cells. Rather than repeat this same selection procedure, it was decided to fill these two cells with paragraphs prepared by sample subjects as a part of the test of the relation between formal logic ability and production of structured writing. While both of these paragraphs treated the same general subject, they contained sufficiently different information to allow independent measurement of comprehension.

These nine paragraphs satisfied all criteria except one. As they were originally written there was a significant relation between readability and percentage of technical terms. To eliminate this significant relation three paragraphs had to be modified by substituting non-technical, but "hard," words for some technical terms. This did not affect the readability levels of these paragraphs, and a retesting of structure showed that the structural levels, also, were unaffected.

It was through these procedures, then, that the nine paragraphs to be tested were selected. The paragraphs are included, here, as Appendix C.
A technical term was defined by a set of criteria developed by this writer for an earlier research into dental vocabulary. It is patterned after a similar definition developed by Denton. A technical term is one that (1) is used only in a medical-dental relation, or (2) has a meaning peculiar to medicine or dentistry, or (3) is characteristic of the activities of medical or dental science and is essential to them even though it is used with the same meaning in other professions, but (4) is not commonly used in a lay-language discussion of the concept.

The borderline drawn by this definition may be illustrated by the words patient and surgeon. Patient means the same thing to both laymen and medical-dental people. Surgeon, however, has a meaning of residency, certification, special attitudes and legal limitations for medical people that is not shared by laymen to whom a surgeon is an M.D. who operates on people. The line is

10National Institutes of Health Grant Number D-1134-A, The Ohio State University Research Foundation Project Number 1068, co-directed by George J. Kienzle and Edgar Dale.

drawn, then, between these two terms. A list of the terms in these nine paragraphs that were classified as technical is included as Appendix E.

The panel of judges who rearranged the sentences to measure structure were six members of the staff of the School of Journalism of The Ohio State University. They measured both the structure of the paragraphs for the nine cells and the structure of the paragraphs used to show the relation between formal logic ability and production of structured writing. This panel was selected for two reasons. First, they represent a group of expert writers. In addition to teaching writing, they all engage in writing for publication. Second, they represent a wide range of different possible approaches to writing. Some are newspaper oriented, one is radio-television oriented, and some are more language oriented. Such a variety could not be found as easily among other groups of expert writers.

A group of judges might have been chosen to represent the normal readers of such material. This might logically have been done for if structure affects comprehensibility it does so by either impeding or improving the ease with which the reader follows the flow of the author's ideas.
A reader, then, not an expert writer, might be in the best position to judge the effect of the flow of ideas. This reason, however, was discarded in choosing expert writers because it is not the effect on a reader that is at issue. What was to be measured was how well the author had tied his paragraph elements together. This is better measured by expert judges than by average readers. Excellent writers are in a better position to judge the bonding and organizational elements which the author used.

Figure 2 contains the vital statistics on the nine selected paragraphs. Readability is expressed by the Dale-Chall raw score and structure is expressed by the Kendall coefficient of concordance obtained from the rearrangements made by the six judges.
Within each of the nine cells, the upper-left quarter is the code designation for the paragraph. The upper-right quarter is the structure score, and the lower-left quarter is the readability score. A high structure score signifies a well-structured paragraph but a high readability score indicates a less readable paragraph.

Table 1 contains information on the other measurable variables of these nine paragraphs.
The two major variables and the three extraneous variables were all analyzed by means of the Wilson analysis of variance. Only two variances were found to be statistically significant. The variance between columns for structure and the between rows variance for readability were both statistically significant. The appropriate chi-squares are:

\[
(1) \chi^2_C (\text{structure}) = 6.30; \text{ df} = 2; \ 0.05 > p > 0.02; \text{ and}
\]

\[
(2) \chi^2_r (\text{readability}) = 6.30; \text{ df} = 2; \ 0.05 > p > 0.02.
\]

<table>
<thead>
<tr>
<th>Paragraph</th>
<th>Sentences</th>
<th>Words</th>
<th>Technical Terms (%)</th>
<th>Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>9</td>
<td>152</td>
<td>13.8</td>
<td>P-P</td>
</tr>
<tr>
<td>M</td>
<td>10</td>
<td>204</td>
<td>5.9</td>
<td>F-P</td>
</tr>
<tr>
<td>O</td>
<td>10</td>
<td>258</td>
<td>2.3</td>
<td>P-F</td>
</tr>
<tr>
<td>R</td>
<td>8</td>
<td>70</td>
<td>20.0</td>
<td>G-P</td>
</tr>
<tr>
<td>N</td>
<td>7</td>
<td>131</td>
<td>6.9</td>
<td>F-F</td>
</tr>
<tr>
<td>X</td>
<td>10</td>
<td>247</td>
<td>13.4</td>
<td>P-G</td>
</tr>
<tr>
<td>D</td>
<td>9</td>
<td>118</td>
<td>6.8</td>
<td>G-F</td>
</tr>
<tr>
<td>A</td>
<td>10</td>
<td>160</td>
<td>6.2</td>
<td>F-G</td>
</tr>
<tr>
<td>E</td>
<td>8</td>
<td>184</td>
<td>7.1</td>
<td>G-G</td>
</tr>
</tbody>
</table>
All other variances, including interactions, had chance occurrences greater than ten per cent. From this it was concluded that these nine paragraphs varied significantly only in structure and readability.

These nine paragraphs will provide useful information only after their comprehensibility has been tested on a sample of readers. This requires sample subjects. The next subsection of this chapter will discuss the selection of these sample subjects.

Three criteria were used in this selection. First, there must be enough subjects to provide for an adequate statistical analysis. Second, the sample subjects must be from the group which normally reads the type of information in the paragraphs. Third, the subjects must be available.

There are two places in this project where sample subjects are necessary, and the number of subjects required is different for each. One phase of the testing calls for measuring the subjects' formal logic ability and for correlating these results with the subjects' ability to write well-structured prose. The proposed procedure requires that this correlation to be made on a sample
all members of which have nearly equal backgrounds in English composition and formal logic training. The number of subjects proposed for use in this correlation is ten.

Ten subjects will not, however, provide an adequate basis for determining the reliability of the formal logic test. Also, the ten subjects were not to be preselected on the basis of equal training, but were to be chosen from a larger convenience sample on the basis of their having had equal training and on the basis of their representing a range of formal logic abilities. The combination of these factors set the minimum number required at about fifty. Fifty subjects would provide for an adequate determination of reliability and ten subjects with nearly equal training backgrounds should be found in a total sample of fifty subjects.

The paragraphs selected were on medically related topics. Consequently the sample subjects must come from a medically related group. To satisfy the availability criterion these subjects were to be gathered from the resident staffs of the University Hospital and Children's Hospital in Columbus, Ohio.
To ease the data handling in the phase of this project where the comprehensibility of the nine paragraphs was to be measured, the number of subjects necessary could be reduced somewhat. In this part of the project eighteen pieces of information are to be collected on each subject. That is, the speed of reading and the comprehension are to be measured on each of nine paragraphs. This is a large amount of data to be treated with nonmachine methods. Still, an adequate index of reliability must be available for each test of comprehension. Thirty-five seemed to be the number of subjects which would be adequate and which would keep the volume of data manageable.

The requirements for availability and professional backgrounds of the subjects were again applicable. And again, the subjects were to be gathered from the resident staffs of the same two hospitals.

Neither the sample sizes nor the selection method allows for much generalization of the findings beyond the samples themselves. It should be restated that the broad goal for this study is to show the feasibility of this method of writing analysis. In order to obtain definitive results that merit generalization it will be necessary to
repeat this project on more subjects selected so that they are representative of a larger population.

These same reasons help explain why this project was done on medical residents and medical writing when the sponsorship of the project is dental-related. While the findings on the specific hypotheses may not be transferred from medicine to dentistry, the feasibility findings may be generalized. Even if a dental sample had been chosen, the findings on the hypotheses could not be generalized beyond the sample. And the medical residents were available during the summer months, when the testing was to be done, while a dental sample was unavailable.

These two sample sizes and sample sources, then, were used as outlined earlier in this methodology chapter.
Chapter III

RESULTS OF THE STUDY

The preceding chapters have been largely a discussion of what was to be done and why it was to be done. This chapter begins the discussion of what was actually done and what was found. This chapter will describe the testing procedures followed, the resulting data, the statistical methods used, and the statistical results relating to the four hypotheses. The interpretation of these results will be the topic for chapter IV.

Formal Logic and Structured Writing

The first part of the proposed study was an examination of the relation between formal logic ability and the ability to write well-structured prose. This requires two kinds of information about one sample of professional writers — an index of formal logic ability and an index of the ability to structure writing. This section will discuss the administration and the results of the measures of these two abilities.
The development of the formal logic instrument was discussed in the preceding chapter and a copy is appended to this report as Appendix A. The sample to be tested was chosen as proposed in the preceding chapter from the resident staffs at University Hospital and Children's Hospital.

The test, itself, consisted of fifteen sets of syllogisms each having one or more conclusions. There was a total of fifty conclusions. The instructions to the subjects were, simply, to mark each conclusion valid or invalid as a conclusion drawn from the given syllogisms.

The conclusions prepared for each set of syllogisms incorporated as many of the rules of formal logic as was possible. That is, those conclusions which were to be correctly marked invalid would have been marked valid by a subject committing the most common fallacies of formal logic. Those conclusions which were to be correctly marked valid included all of the possible valid conclusions that could be properly drawn from the given syllogisms. The results seemed to indicate that the problems became more difficult toward the end of the test although there was no attempt to prepare the test in this way.
The subjects had five minutes in which to attempt all fifty conclusions. An informal pretesting of the instrument on eight dental graduate students had shown that the average time to complete the test was about ten minutes. Five minutes was chosen as the time limit in order to center the distribution of scores near the fifty per cent level. In this way, nearly all the scores could be kept below the 100 per cent mark. This would avoid a skewing of the distribution by the concentration of many scores at the possible maximum.

The test was administered to 27 surgical residents at University Hospital and 29 pediatric residents at Children's Hospital. These groups were tested in two separate administrations. The instructions to each group were identical and were exactly as printed on the test papers. No questions were raised by the subjects.

Testing in two separate sessions, even with standardized instructions, allows the possibility of biasing error. These possible errors, however, are considered insignificant in this report and they will not be treated further.
Figure 3 shows the distribution of scores on this test.

Figure 8

Distribution of Scores on the Test of Formal Logic Ability
This figure shows that the range of scores was from 28 per cent to 90 per cent. The median was 56 per cent. There was a slight pile-up at the lower end of the distribution because the sample contained three residents who are not native speakers of English. Apparently their problems of translation acted to reduce their scores. Their papers were eliminated from further consideration.

The reliability of this test was measured by the split-half technique. The results on the odd numbered items were correlated with the results on the even numbered items by means of the Spearmen rank-order correlation corrected for ties.¹ The coefficient of reliability was not corrected for the fact that the split-half technique reduces the size of the correlated tests by one half. The coefficient of reliability thus obtained was 0.85. The probability of achieving an equally high coefficient by chance alone is less than 0.1 per cent. The measure, then, has a high and significant reliability.

The validity of the instrument was not estimated statistically. The test does have logical validity in that the required task is one of the overt behaviors resulting from formal logic ability. The test is similar to other tests of this same ability, and this test was constructed to cover most of the rules which apply to this type of behavior. There is no claim made for validity beyond this.

The results on this measure were to be correlated with the subjects' ability to write well-structured prose. The measure of this second ability was obtained by having each subject write a paragraph, dividing these paragraphs into sentences, and measuring the agreement of the panel of judges rearranging the sentences.

Each subject was asked to write a paragraph defending the practice of the surgical scrub. All would have more than enough factual material on which to base such an argument. The instructions called for one cohesive paragraph of from eight to twelve sentences. These limits on length were necessary in order to control this variable. The chances of the judges having perfect agreement on the order of three sentences, for example, are much greater than the chances of achieving perfect agreement on ten
sentences. The actual instructions given to the subjects are included as a part of Appendix B.

All subject-written paragraphs were not to be submitted to the judges. Ten of these paragraphs were to be selected so that (1) the subjects' backgrounds in formal logic training and English composition were about equal, and (2) the subjects represented the whole range of formal logic ability scores. To accomplish this, the formal logic ability distribution was divided into deciles. From each decile one subject was selected. Backgrounds in the two training variables were correlated against formal logic ability scores. Various combinations were tried until ten were found which had zero correlations between the training variables and formal logic scores.

No subject was chosen from the lowest decile. Three of the four subjects in this decile were foreign students and the fourth wrote only a four-sentence paragraph. Nine paragraphs, then, were submitted to the panel of judges.

The measure of the degree of structure of the paragraphs was the extent to which the judges agreed among themselves on what the order of sentences should be. Kendall's coefficient of concordance was used as the
statistical index of this agreement. This statistic is, in a sense, an average of all possible correlations between pairs of orders. The coefficient can vary from 0 to 1 and the significance of the coefficient can be determined.

The reliability of the panel of judges was determined by using the split-half technique and the Spearman rank-order coefficient of correlation. This reliability was based on these nine paragraphs plus the eight paragraphs submitted to them in preparation for placing the paragraphs into matrix cells. The reliability of the judges was 0.75. This correlation is statistically significant beyond the 1 per cent level.

Table 2 gives the specific results on formal logic and structure by the ten selected subjects.
The Spearman rank-order correlation between the two variables is 0.62. This correlation is significant beyond the 5 per cent level (.05 > p > .02).

The decision, then, is to reject the null hypothesis and conclude that there is a significant relationship between formal logic ability and ability to structure writing. This conclusion is drawn, of course, within the limits of experimental error already discussed.
Structure, Readability and Comprehensibility

The remaining three hypotheses of this study were tested simultaneously by the second half of this investigation. The selection of the nine paragraphs and the designing of the nine comprehension tests were discussed in the preceding chapter.

Table 3 gives a complete listing of data pertinent to these nine paragraphs.

Table 3

Complete Data on the Nine Paragraphs

<table>
<thead>
<tr>
<th>(1) Paragraph Code Letters</th>
<th>(2) Dale-Chall raw scores</th>
<th>(3) Structure Scores</th>
<th>(4) Matrix Cells</th>
<th>(5) No. of Words</th>
<th>(6) No. of Sentences</th>
<th>(7) % of Technical Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>9.88</td>
<td>0.33</td>
<td>P-P</td>
<td>152</td>
<td>9</td>
<td>13.8</td>
</tr>
<tr>
<td>O</td>
<td>9.87</td>
<td>0.51</td>
<td>P-F</td>
<td>258</td>
<td>10</td>
<td>2.3</td>
</tr>
<tr>
<td>X</td>
<td>9.92</td>
<td>0.83</td>
<td>P-G</td>
<td>247</td>
<td>10</td>
<td>13.4</td>
</tr>
<tr>
<td>M</td>
<td>8.97</td>
<td>0.34</td>
<td>F-P</td>
<td>204</td>
<td>10</td>
<td>5.9</td>
</tr>
<tr>
<td>N</td>
<td>8.89</td>
<td>0.53</td>
<td>F-F</td>
<td>131</td>
<td>7</td>
<td>6.9</td>
</tr>
<tr>
<td>A</td>
<td>8.77</td>
<td>0.80</td>
<td>F-G</td>
<td>160</td>
<td>10</td>
<td>6.2</td>
</tr>
<tr>
<td>R</td>
<td>7.69</td>
<td>0.35</td>
<td>G-P</td>
<td>70</td>
<td>8</td>
<td>20.0</td>
</tr>
<tr>
<td>D</td>
<td>7.49</td>
<td>0.51</td>
<td>G-F</td>
<td>118</td>
<td>9</td>
<td>6.8</td>
</tr>
<tr>
<td>E</td>
<td>7.26</td>
<td>0.84</td>
<td>G-G</td>
<td>184</td>
<td>8</td>
<td>7.1</td>
</tr>
</tbody>
</table>
Column 2, the Dale-Chall Readability Formula raw scores, are just that -- raw scores. They were not converted to school grade levels because there was no need to do so. Actually the raw scores in the interval 9.00 - 9.99 are equivalent to school grades 13-15. That is, college material. The interval 8.00-8.99 is equivalent to grades 11-12, and the interval 7.00-7.99 is equivalent to grades 9-10.

It can be seen from this column that paragraphs U, O, and X are the three which make up the poor readability matrix row. Paragraphs M, N, and A are medium or fair readability; and paragraphs R, D, and E have a good readability level.

Column 3 gives the structure scores of the nine paragraphs. These are the Kendall coefficients of concordance of the amount of agreement on sequence among the panel of six judges. It can be seen from this column that paragraphs U, M, and R have poor degrees of structure; paragraphs O, N, and D have fair structure; and paragraphs X, A, and E have good structure.
Column 4 just summarizes the information which was presented in the preceding paragraphs of this discussion. The letters P, F, and G, obviously stand for poor, fair and good. The pair of letters in the column has readability first and structure last. Thus, P-F means poor readability and fair structure.

Columns 5 and 6 give the total number of words and sentences in each paragraph. No special rules were used in these countings. A word is, simply, a group of letters set off from other groups by a space. A sentence is, simply, a group of words followed by a period, question mark or exclamation point.

The percentage of technical words, column 7, is the number of technical terms in the paragraph multiplied by 100 and divided by the total number of words in the paragraph. The definition of a technical term was presented in the preceding chapter. A list of the actual terms from the paragraphs which were classified as technical is appended to this report as Appendix E.

Table 4 gives the results of Wilson nonparametric analyses of variance on all measured variables of the nine paragraphs.
Table 4

Nonparametric Analyses of Variance on Measured Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\chi^2_t$</th>
<th>df</th>
<th>$\chi^2_r$</th>
<th>df</th>
<th>$\chi^2_c$</th>
<th>df</th>
<th>$\chi^2_I$</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Readability</td>
<td>9.0 8</td>
<td>6.30* 2</td>
<td>0.90 2</td>
<td>1.80 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure</td>
<td>9.0 8</td>
<td>0.90 2</td>
<td>6.30* 2</td>
<td>1.80 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No.of Words</td>
<td>9.0 8</td>
<td>0.90 2</td>
<td>7.20 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No.of Sentences</td>
<td>9.0 8</td>
<td>3.60 2</td>
<td>4.50 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of Technical</td>
<td>9.0 8</td>
<td>3.60 2</td>
<td>1.80 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*S*Significant beyond the 5% level (*0.05 > P > 0.02*).

Only two variances were significant beyond the five per cent level. The between-rows variance for readability and the between-columns variance for structure varied significantly in the intended directions. All other variables were considered to have insignificant arrangement patterns.

One other variable has not, as yet, been mentioned. This is the subject matter of the paragraphs. Subject matter could not be quantified, but the following table shows the variety of subjects included.
Table 5

Subject Matter of the Paragraphs

<table>
<thead>
<tr>
<th>Paragraph Code No.</th>
<th>Matrix Cell</th>
<th>Subject Matter</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>P-P</td>
<td>fibroid polyps</td>
</tr>
<tr>
<td>O</td>
<td>P-F</td>
<td>qualities of surgeons</td>
</tr>
<tr>
<td>X</td>
<td>P-G</td>
<td>measuring blood pH</td>
</tr>
<tr>
<td>M</td>
<td>F-P</td>
<td>diagnosis of hyperparathyroidism</td>
</tr>
<tr>
<td>N</td>
<td>F-F</td>
<td>basis for surgical scrub</td>
</tr>
<tr>
<td>A</td>
<td>F-B</td>
<td>treatment of gangrene</td>
</tr>
<tr>
<td>R</td>
<td>G-P</td>
<td>antisepsis and healing</td>
</tr>
<tr>
<td>D</td>
<td>G-F</td>
<td>abdominal injuries</td>
</tr>
<tr>
<td>E</td>
<td>G-G</td>
<td>the calorie</td>
</tr>
</tbody>
</table>

This list should show that there is no obviously apparent pattern to the arrangement of topics in the matrix.

Comprehension tests were written to cover the information contained in each paragraph. These nine tests varied in length from 10 to 15 questions. Each test was designed so that most significant items of information in the paragraph were tested.
All items were of the true-false type. But to help overcome the effect of prior knowledge on the responses, an additional criterion of trueness was added. An item in order to be true must be both an accurate statement and be based on information contained in the paragraph. A false item might be either an inaccurate paraphrase of a statement in the paragraph or a true statement of fact not covered in the paragraph. In the test of the paragraph dealing with blood pH, for example, one item said that pOH was an index of the concentration of hydroxyl ions. This statement is an accurate statement of what pOH is, but the paragraph dealt only with pH. Therefore the correct response to this item was false.

Copies of the nine tests used may be found at the end of this report as Appendix D.

There was no limit placed on the amount of time a subject could take to respond to these tests. Once a subject began to take a test, however, he was not permitted to reread the paragraph.

The nine paragraphs and the nine tests were administered to 20 residents in general medicine at University Hospital and 16 pediatric residents at Children's Hospital.
Each group was tested separately. Again, these separate administrations admit the possibility of certain experimental errors, but, again, they were not considered significant.

The paragraphs and tests were arranged into one test booklet. The paragraph to be read was always followed by the corresponding test, but the paragraphs were arranged in a different order in each booklet. This changing of the order helped control such experimental errors as practice and fatigue by randomizing such variables throughout the design.

The instructions to the subjects were standardized by reading only the instructions printed on the test booklet. The subjects raised no additional questions. The actual instructions are appended with the paragraphs in Appendix C. They were, briefly, to read each paragraph, record the time spent reading, answer the questions, and repeat the procedure for the remaining eight paragraphs.

Each subject timed himself. It was correctly believed when this procedure was planned that a physician without a watch with a sweep second hand would be a rarity. Of the 36, only one was not so equipped and he was supplied with such a watch.
Eighteen scores were computed for each sample subject. Reading rate in words per second and comprehension in per cent of test items correct for each of the nine paragraphs were computed.

The test results from three subjects were not analyzed because the residents were not native speakers of English. This left a total of 33 usable sets of data.

The two scores on each paragraph were combined to give one index of the comprehensibility of the paragraph for each subject. This was done by simply multiplying the two scores. That is, the index of the comprehensibility of a paragraph for a subject was the simple product of his reading rate on the paragraph and the percentage of test items he answered correctly. For example, a paragraph read at 4.5 words per second on which the comprehension was 80 per cent would have a comprehensibility index of 360.

This procedure contains one other possibility of experimental error. The nine tests may not have been of equal difficulty. The variance of test scores may, then, reflect the varying difficulty of the tests and not the varying comprehensibility of the paragraphs. This possible error was discounted here, though, for three reasons. One,
the paragraphs were rather similar in over-all style and content and there seemed to be no reason to believe that any particular paragraph would lend itself to a more difficult test than any other paragraph. Two, all of the tests were written by the same person at the same time. This would probably cause the tests to be of similar difficulty. And three, since there was no obvious pattern to the arrangement of topics in the design it would seem to follow that there would be no pattern to varying test difficulty in the design.

Figure 4 shows the median scores for reading rate and percentage of comprehension for the nine paragraphs.

Figure 4

The Nine Cells and Test Results

<table>
<thead>
<tr>
<th>Readability</th>
<th>Structure</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>4.08</td>
<td>3.74</td>
<td>3.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>86.7</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td>3.04</td>
<td>3.22</td>
<td>4.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>80</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>2.80</td>
<td>3.37</td>
<td>4.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>80</td>
<td>72.7</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>Fair</td>
<td>Good</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In Figure 4, the upper number in each cell is the median reading rate for that paragraph in words per second. The lower number is the median percentage of correct responses. The comprehension percentages in the poor readability-good structure cell and the good readability-fair structure cell are not multiples of 10 because these two tests contained more than 10 items. All other tests had exactly ten items.

These nine tests had varying reliabilities which were determined by using a combination of the split-half technique and the Spearman rank-order correlation. While all of the reliability coefficients are statistically significant, most of them are not high enough to be classified as excellent.

It may be that the tests are more reliable than the coefficients indicate because the number of subjects was small. The fact that the results of the entire experiment were in accord with the hypotheses gives some credence to this possibility.

Table 6 lists the coefficients of reliability of the nine comprehension tests. Here again the readability-structure designation has been used to describe the cells.
Table 6

Reliabilities of the Comprehension Tests

<table>
<thead>
<tr>
<th>Paragraph</th>
<th>Cell</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>G-P</td>
<td>0.30</td>
</tr>
<tr>
<td>M</td>
<td>F-P</td>
<td>0.70</td>
</tr>
<tr>
<td>U</td>
<td>P-P</td>
<td>0.83</td>
</tr>
<tr>
<td>D</td>
<td>G-F</td>
<td>0.45</td>
</tr>
<tr>
<td>N</td>
<td>F-F</td>
<td>0.49</td>
</tr>
<tr>
<td>O</td>
<td>F-P</td>
<td>0.40</td>
</tr>
<tr>
<td>E</td>
<td>G-G</td>
<td>0.68</td>
</tr>
<tr>
<td>A</td>
<td>F-G</td>
<td>0.71</td>
</tr>
<tr>
<td>X</td>
<td>P-G</td>
<td>0.50</td>
</tr>
</tbody>
</table>

It does not follow that the median comprehensibility indices made from the products of rate and comprehension will be equal to the product of the medians of these scores. Therefore, in Figure 5, which contains the median comprehensibility indices for the nine paragraphs, the figures are not necessarily equal to the products of the appropriate two numbers in Figure 4.
The data in this figure are the final data used to test the remaining three hypotheses of this investigation.

Hypothesis 2 stated that comprehensibility would vary significantly with readability. An analysis of the final comprehensibility scores showed that the between-rows variance was significant beyond the one per cent level.

Hypothesis 3 stated that comprehensibility would vary significantly with levels of structure. The between-columns variance which tested this hypothesis was significant beyond the one per cent level.
Hypothesis 4 was that these two variables, readability and structure, would act independently in their effect on comprehensibility. The interaction variance was not statistically significant. This insignificance tends to support the hypothesis.

These statistical analyses employed the Wilson non-parametric analysis of variance. This technique is a modification of the Extension of the Median Test which is, in turn, an extension of the chi-square technique. This method was chosen for this analysis because it was the only nonparametric statistical procedure found which provided a means of determining the total variance in the matrix. Without this estimation of total variance, the interaction effect, if any, can not be measured and the lack of a significant interaction effect was one of the hypotheses of this study.

Like the Extension of the Median Test, the Wilson technique begins by selecting a score (an approximate median) which divides all of the matrix data into two, about equal parts. The number of cases in each cell which

\[ \text{Number of cases in each cell which} \]

\[ ^2 \text{Wilson, op. cit.} \]

\[ ^3 \text{Seigel, op. cit., pp. 179-184.} \]
fall above this number and the number of cases in each cell below this number provide the basis for the statistical analysis. The approximate median of all comprehensibility indices was 275. Figure 6, below, gives the number of cases which are above and below this score in each cell.

**Figure 6**

**Number of Observations Above and Below Median**

<table>
<thead>
<tr>
<th>Above Md.</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>20</td>
<td>21</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>22</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>18</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>46</td>
<td>61</td>
<td>146</td>
<td></td>
</tr>
<tr>
<td>Below Md.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>12</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>21</td>
<td>11</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>19</td>
<td>15</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>53</td>
<td>38</td>
<td>151</td>
<td></td>
</tr>
</tbody>
</table>

Applying the Wilson technique to this data gives the following results:

1) \( \chi^2_T \) (total variance) = 25.9; \( df = 8 \)
   \[ .01 > p > .001 \]

2) \( \chi^2_R \) (readability) = 10.2; \( df = 2 \)
   \[ .01 > p > .001 \]

3) \( \chi^2_C \) (structure) = 10.2; \( df = 2 \)
   \[ .01 > p > .001 \]

4) \( \chi^2_I \) (interaction) = 5.5; \( df = 4 \)
   \[ .30 > p > .20 \]
Summary. This chapter has treated the actual experimental procedures followed and the results obtained. Hypothesis 1 was tested against 9 cases selected from a total usable sample of 53 medical residents. The results showed a correlation of 0.62 between formal logic ability and ability to write structured prose. This relation was high and statistically significant. The results tended to support the hypothesis.

Hypotheses 2, 3 and 4 were tested against a usable sample of 33 medical residents. The results showed that comprehensibility of paragraphs varies significantly with different levels of readability (Hypothesis 2), and with different levels of structure (Hypothesis 3). The results also showed no statistically significant interaction between readability and structure (Hypothesis 4). The experimental results tended to support, then, each of these three hypotheses.

In the discussion of actual procedures, there was some treatment of the sources of possible experimental errors.
The following chapter of this report will deal with the conclusions drawn from this study. It will cover conclusions, implications, and areas for further study which relate to the four hypotheses and the major question of the feasibility of this method of writing analysis.
Chapter IV

CONCLUSIONS AND IMPLICATIONS

This chapter will deal with the conclusions which may be drawn from the work discussed in the preceding chapters. It will include some commentary on the limits within which these conclusions may be validly drawn. The chapter will also treat the implications which this work holds for writing in general and professional writing in particular. There will also be some discussion of areas which were opened for study or were left to be studied later.

Conclusions and Their Implications

The results of this study have two aspects. First, there is the question of the feasibility of this type of analysis of writing. Second, there are the four specific hypotheses which were tested.

Feasible, here, means something more than simply possible. It has the additional connotation of suitable or appropriate. Further, it carries the connotation of success. In this study the question of feasibility
requires an answer which includes answers to subquestions like: "Is it possible to use the rearrangement test to measure structure?" "How successful is this test?" "Is this test economical, or is there a better, easier test already available?" and "Even if it is successful, how important is it?"

As a result of this project at least partial answers to these questions are available.

There is reasonably good evidence to support the belief that there is a structural element in writing. This element has been discussed theoretically by showing how it is possible to have different kinds of structure and to have both effective and noneffective structure.

An element called structure has been isolated experimentally in this investigation. The development of the nine test paragraphs showed that material could be written which had levels of this element which did not relate to any commonly measured aspect of writing. There was no significant correlation between this element and length of the paragraph, number of sentences, average length of sentences, difficult words, technical words, or formula readability.
This operationally defined structure was shown to have an effect on the comprehensibility of writing. This effect was shown to be independent of the effect of readability.

An important question remains as to whether or not the operationally defined structure is the same element as the theoretically defined structure. There seems to be logical evidence for believing that they are identical. The rearrangement test measures the strength of the bonding between paragraph information. That is, two pieces of information, two sentences, which had a strong linkage would be most likely to be put back together by judges. Structure, defined theoretically is, again, this same bonding of information.

Analyzing the structural element of writing in this way is feasible. The rearrangement test has a logical validity and it is as reliable as a panel of judges can be. This element of writing is not necessarily related to any other element of writing and it seems to be important for comprehensibility.
The rearrangement test is neither as economical nor as objective as it might be. Using a panel of judges to reorder the sentence particles is both subjective and time-consuming. The purpose of this project was, however, to examine this test, not to try to find a better test. This study did show, however, that in order to find a faster, more objective method it will be necessary to measure elements of writing other than those studied here.

The conclusion that this method is feasible must be modified somewhat because of the design and the samples used. The meanings of feasible which have possible and appropriate as synonyms can be applied without limitation to this method because appropriateness and possibility are matters relating to the validity of the method. The logical validity shown for this structural analysis is not affected by either the design of this investigation or by the samples used.

The meaning of feasible related to success, however, is affected by design and samples. There were but nine paragraphs studied and these were tested against only 33 subjects. It is not possible to generalize these findings much beyond the actual samples themselves. This study
showed that structure was not necessarily related to readability, for example, but it may be that they are normally related. It would seem quite reasonable to believe that excellent writing is both highly readable and well structured while poor writing is both less readable and more poorly structured. It is possible, too, that on different paragraphs, different tests of comprehension, and different subjects structure would not have a significant, independent effect on comprehensibility.

It can be safely concluded, though, that under the conditions of this investigation this analysis of structure was successful. There is a possibility that it may not be successful under other conditions, but this is only a possibility and would have to be demonstrated.

This analysis of structure, then, is feasible in that it is possible; it is appropriate; and it is, under the conditions of this study, successful.

The conclusions drawn regarding the four specific hypotheses have been briefly mentioned in the discussion of feasibility. And the limits placed on such conclusions by the number of paragraphs tested and the number of sample subjects have been mentioned. But there are
further limitations placed on conclusions relating specifically to the four hypotheses by other experimental procedures.

The first hypothesis stated that a relation exists between an individual's ability in formal logic tasks and his ability to structure his writing. This was tested by administering a test requiring the subject to manipulate symbols according to the rules of formal logic to about 50 medical residents. Paragraphs written by nine of these subjects whose scores encompassed nearly the entire range of formal logic test scores were studied to determine the degree of the structure element. And, indeed, there was a high correlation between the two abilities.

There remain, however, critical questions as to the validity of the two ability measures. Both tests seem, logically, to measure the two abilities, but no experimental evidence can be offered in support of this contention. Also, the subjects wrote a specified amount on a specified topic to develop a specified kind of argument. This may have induced an error-causing artificiality into the situation. No evidence is available to support
generalizing the result to any population larger than the nine sample cases.

With these limitations, what may be validly concluded about this hypothesis? Just this: under the conditions used in this study and on the given sample, the hypothesis was supported.

There is a truism of logic that it takes an infinitely large number of tests to prove something conclusively, but it takes only one unsuccessful trial to disprove something. To this extent, then, this hypothesis has been proved. This is not the case which disproves the hypothesis and there is now some evidence which supports it.

Nine paragraphs were prepared so that when they were arranged in a 3x3 matrix they had three levels of readability (rows) and three levels of structure (columns). The paragraphs were prepared so that these were the only two variables which had significant variances in the two primary dimensions of the matrix. Nine is a rather small number of paragraphs to represent writing generally, professional writing more specifically, or medical-dental writing even more specifically.
All of the paragraphs were similar in topic, length, number of sentences, and technical vocabulary. While these similarities were controlling extraneous variables, they may, at the same time, have been providing an error-causing artificiality. The results achieved may only hold for the levels of these extraneous variables which were used in this study.

The comprehensibility of these paragraphs was tested on only about thirty medical residents. No support can be found for generalizing the findings to any larger population.

The indices of comprehensibility of the paragraphs were measured by comprehension tests and by the rate at which the paragraphs were read. The tests of comprehension had only fair reliabilities; these tests may have been of unequal difficulty; and the rate of reading may have been influenced by the length of the passages. There might, then, be some question raised as to the reliability of these indices.

With what assurance, then, can it be said that structure affects comprehensibility independent of readability? As with the first hypothesis, it may safely be
concluded that these three hypotheses are not proved; they are supported; and they are not disproved. Under the experimental conditions used, the hypotheses do hold.

In summary, then, these are the conclusions which may be drawn from this study.

1. There is an element of writing which gives writing a unity and flow of information. This element may be called structure.

2. The rearrangement test has a limited feasibility in measuring this element. The test does measure a unique quality of writing and there is some support for believing that this quality is structure.

3. There is some evidence to support an hypothesis that under certain conditions, at least, the ability to structure writing is related to the ability to apply the rules of formal logic to verbal syllogisms.

4. There is some evidence to support an hypothesis that under certain conditions structure affects the comprehensibility of written material.
5. There is additional evidence to support the already tested hypothesis that readability affects the comprehensibility of writing.

6. There is some evidence to support an hypothesis that under the conditions used in this study readability and structure act independently in their effect on comprehensibility.

**Implications of the Study**

An implication is an inference which may be drawn on the basis of information at hand. To discuss the implications of this study would be to discuss all of the logically consistent inferences which might be drawn from this investigation. But such a discussion would be rather undirected. Consequently, this section will be limited to the implications which this study has for (1) writing generally and (2) medical-dental writing specifically. This is necessarily an arbitrary limitation of the discussion, but it still includes the two areas where this study has its most direct impact.
The discussion of the implications of this study for each of these two areas will have, rather, three parts. The first of these will be the implications of the conclusion that this method of analysis is feasible. The other two parts will center about the alternative conclusions relating to the four hypotheses. That is, both the possibility that the hypotheses hold for situations broader than the experimental situation and the possibility that the hypotheses would not hold in any broader situation will be examined.

The rearrangement test seems to measure a quality of writing that has not been previously quantified. What does this mean for writing generally?

It would seem that structure is, now, a stylistic element of writing. It would, further, seem that this element is measurable. In light of the arguments presented in the first chapter of this report, it would seem that structure is more mechanical than stylistic. That is, those elements of writing which may be defined by some pattern, some system of general rules, are not exclusively products of creativity, of art, or of talent, but are elements which may possibly be taught to all students of writing.
In much the same way that the production of readable writing is becoming a widely teachable skill, it would seem that the achievement of structure might become a widely teachable skill. This implies two ideas. First, it implies that teaching students to structure their writing may be simply teaching them the certain generalizations by which all writing is structured. Second, it implies that there may be aspects of writing now included in style which may be quantified and thus be taught to all who must write.

Nothing here implies that great writing results simply from following the rules or patterns for composition. What is intended by this discussion is to imply that structure may become another useful tool in the attempt to produce many adequate writers. Training such writers is much easier when there are general guidelines which may be followed.

This study did not show how writing is structured. Nor did it show what causes some structures to be more easily comprehended than others. The study did show that, within limits, the rearrangement test is capable of measuring structure. Thus all of the implications of this study are written only as possibilities.
The method of measuring structure has its greatest implications for professional writing such as medical and dental journalism. These areas are much more concerned with communicating information efficiently than with producing lasting works of verbal art. The aim of instruction in writing for people in these areas is more concerned with training adequate writers than with producing great writers. Consequently, any system which reduces the time required to achieve adequacy of writing will have a pronounced effect on instruction. The system should, as well, have a positive effect on the quality of the writing in these areas because adequate writing will be easier to achieve by both writers and editors.

The rearrangement test, then, has implications for the whole field of writing. The effects of those implications, however, should be greatest and be felt first in professional writing such as medical and dental journalism.

The implications which the four hypotheses have for the fields of writing and professional writing must be expressed as alternatives. The results of this study are more indications than proof. Consequently, implications must be either: "If the hypothesis is true, then...." or
"If the hypothesis is not true, then...." This is not an evasion of responsibility in drawing implications, but only a reasonable consideration of the actual results of this investigation.

If the hypotheses are true for situations broader than the experimental one, then the following implications should hold.

1. Some people are, as a result of an as yet undescribed kind of training, best able to structure their writing. Training in formal logic would seem to be one of the ways of improving this skill. The relation between these two abilities, established by this study, was only correlational and not causal. This leaves three possibilities. Either improving logic skill will improve structure skill, and vice versa, or both will be improved by training in a third, unknown skill. This third possibility does not seem too likely because of the great similarity of the tasks which require the two abilities. What appears most reasonable is to believe that training in either ability would improve the other.
2. Structure is a significant element in communication. This would imply that there are particular structural patterns which are easier for a reader to follow or which tend to be more expected by a reader. This would mean that reading is easier when the reader finds the pattern he expects or when the pattern is easy to follow. Reading would be difficult when the flow of ideas is difficult to follow or when the order is one that the reader did not expect to see used.

3. Structure is uniquely different from all other quantified elements of writing. This would mean that predictions of comprehensibility now made by estimating readability could be improved by including an estimation of structure. It also would mean that no quality of writing which we have tried, so far, to measure objectively will estimate structure. These implications might have their greatest effect in the area of professional writing. If training in formal logic improves the ability to structure writing, then courses which improve the logical skill of the students would also improve their ability to write. Or the reverse,
proper writing instruction would improve students' skill in the logic of their professions. It may be that writing instruction which includes training in how to structure writing would improve the research thinking of physicians and dentists. Such a possibility would have a tremendous effect on professional education.

It might be that the training which physicians and dentists receive already equips them to structure writing well. Or it may be that the natural process of selection undergone by these people makes them, as a group, the best equipped people to write well-structured prose. These, of course, are but possibilities, but they are not so remote that they may be ignored.

If, on the other hand, the hypotheses are not true, then the following implications may be drawn.

1. Structuring writing is not related to formal logic ability. This would mean that people structure writing regardless of their ability to apply the rules of formal logic. Thus, training in logic would not improve a person's skill as a writer nor would writing training improve a person's logic ability.
2. Structure is not a significant element in communication. This would mean that there would be no relation between the way ideas are arranged and the ease with which readers follow the sequence of ideas and derive meaning from the ideas.

3. Structure is not an independent quality of writing. It must, then, be related so closely to another measurable quality of writing that measuring the effect of one also measures the effect of the other. We would need not estimate the structural level of writing in order to predict its comprehensibility.

The evidence gathered by this study does not conclusively deny these last three implications. But the evidence does tend to indicate that these last three are not the proper implications of this study.
Directions for Future Investigation

The directions which future study might take are almost limitless because this project touched on so many areas. It has pointed up the need for answers to many more questions than it sought answers for. This study has accomplished two things in this regard. First, it has shown a few new areas where further research might be fruitfully done, and second, it has pointed up the almost complete lack of research in many other areas.

The search for background for this study showed the meager amount of experimental research devoted to the writing process. The nature of the activity as it is performed by an author is still a sort of mystical process by which thoughts are made comprehensible to others. A prospective writer can find a few guides to follow, but he is left to try "trial-and-error" to find most other patterns for his communication.

The mental aspects of writing are almost virgin research territory. No one has explained the similarities or the differences between the mental aspects of the various modes of verbal behavior. No one has done much
in explaining the different behavior patterns involved in receiving and transmitting written messages. Even simple questions remain unanswered. For instance, "Is a good reader necessarily a good writer?"

Writing is still, apparently, considered an art form that does not lend itself to quantifiable study. The study of the concept of readability opened the process of writing to quantitative study, but a great deal of the process remains even untheorized.

This project did not open these areas for study. They were open for investigation long before this study was conceived. But they are open and they still should be studied. This project has shown, in a small way, that writing is capable of further, quantitative study.

In the area of medical-dental writing, this study has raised some rather specific questions. For example, "Are science-oriented people the best qualified to structure their writing?"

If structuring writing is related to logical ordering of reason, then there is a possibility that there should be two aspects to writing instruction. Students
would be taught both the skill of ordering and the skills of grammar, punctuation, etc.

It would appear that as a concomitant learning of all education some people have already learned to structure their writing. Others have not acquired this skill. It would also appear that instruction in ordering would be more difficult than instruction in the other writing skills. This might justify dividing basic writing courses into these two types of instruction.

But all of these are possibilities which require much further study before they might be attempted.

In summary, then, it might be said that this project has shown that further quantitative research of writing is possible. It has opened a few new areas to study and it has pointed up the need for a great deal of research on writing. But most important, for the intended purposes of the study, it has shown that the rearrangement test is a feasible method for estimating an important quality of writing that had not before been measured.
Appendix A

A TEST OF LOGICAL REASONING ABILITY

Directions: This is a test of your ability to reason logically. Each test item contains two or more premises and several conclusions. You are to assume that all premises are true even if they disagree with facts you know to be true. You are to indicate, by writing a "V" (for valid) or an "I" (for invalid) in the space beside each conclusion, whether or not that conclusion may be validly drawn from the stated premises.

There are 15 items in the test. Some items may have no valid conclusions; some may have more than one. You will have 5 minutes in which to work. Do not begin until directed to do so, and come to a complete stop when directed.

Example:

I. Premises: A) All men are mortal.
   B) My professor is a man.

Conclusions:

_______ 1) My professor is mortal.
_______ 2) My professor is human.
_______ 3) All mortals are men.
_______ 4) All professors are mortal.
_______ 5) All professors are men.

Discussion: (In this example, the only conclusion that may validly be drawn from the premises is that, "My professor is mortal." Therefore, conclusion 1 should be marked with a "V" and all other conclusions marked with an "I.""

DO NOT TURN THE PAGE UNTIL DIRECTED TO DO SO.
I. Premises: A) All trout are mammals.  
B) All mammals have hair.

Conclusions:

1) All trout are aquatic.
2) Fish are not mammals.
3) All trout have hair.
4) All mammals are trout.
5) All hairy mammals are trout.

II. Premises: A) Some birds have long tails. 
B) All dogs have long tails.

Conclusions:

6) All dogs are birds.
7) Some birds are dogs.
8) Some birds are not dogs.
9) All long-tailed animals are dogs.
10) Some long tails belong to birds.

III. Premises: A) Susan is a pretty girl. 
B) Some pretty girls are not married.

Conclusions:

11) Susan is not married.
12) Susan is married.
IV. Premises: A) My hat is either brown or gray.  
B) My hat is not gray.  

Conclusions:  
_______ 13) My hat is not brown.  
_______ 14) My hat is gray.  
_______ 15) My hat is brown.  

V. Premises: A) None but the pure are secure.  
B) No student is secure.  

Conclusions:  
_______ 16) All students are pure.  
_______ 17) Some pure people are not secure.  
_______ 18) All students are not pure.  
_______ 19) No student is pure.  

VI. Premises: A) Snakes and lizards are reptiles.  
B) Reptiles and birds are not oviparous.  

Conclusions:  
_______ 20) Snakes are not oviparous.  
_______ 21) Birds are reptiles.  
_______ 22) Lizards are oviparous.  
_______ 23) Some snakes are birds.  
_______ 24) Some reptiles are snakes.
VII. Premises:  
A) All teachers are men.  
B) Some men are stupid.  

Conclusions:  
25) All teachers are stupid.  
26) Some teachers are stupid.  
27) No teacher is stupid.  

VIII. Premises:  
A) I am neither tall nor short.  
B) I am not tall.  

Conclusions:  
28) I am short.  

IX. Premises:  
A) Bob likes Bill.  
B) Whoever likes Bill likes Bert.  
C) Bob likes only brawny boys.  

Conclusions:  
29) Bert is brawny.  
30) Bill is brawny.  
31) Bob likes Bert.
X. Premises: A) All men are mortal.  
B) Some mortals are not men.  
C) Robert is not a man.

Conclusions:

_______ 32) Robert is a mortal.
_______ 33) Not all men are mortal.
_______ 34) Robert is either mortal or not mortal.

XI. Premises: A) All Presidents were politicians.  
B) Hoover was an engineer.  
C) Men can have only one profession.

Conclusions:

_______ 35) Hoover was not President.
_______ 36) Hoover was both President and an engineer.
_______ 37) Hoover was not a politician.

XII. Premises: A) All furniture is manufactured.  
B) All chairs are seats.  
C) All seats are manufactured.

Conclusions:

_______ 38) Chairs are furniture.
_______ 39) Chairs are manufactured.
_______ 40) All seats are chairs.
XIII. Premises:  
A) John loves Mary.  
B) Mary loves Jim.  
C) Jim loves Martha.  
D) Martha loves John.  

Conclusions:  
_______ 41) John loves Martha.  
_______ 42) Mary loves John.  
_______ 43) Jim loves Mary.  
_______ 44) Martha loves Jim.  

XIV. Premises:  
A) Missing articles are either stolen or lost.  
B) My watch is missing.  
C) Harry is a sneak thief.  

Conclusions:  
_______ 45) Harry stole my watch.  
_______ 46) My watch was lost.  
_______ 47) My watch is either lost or Harry stole it.  
_______ 48) Either Harry or someone else stole my watch or else it was lost.
XV. Premises:  A) If I were President, I would be worried.
               B) I am not President.

Conclusions:

49) I am not worried.

50) Only worried people are President.
Paragraph 1 -- Structure Score = 0.71 -- Logic Score = 68

It is a known fact that bacteria, including pathogenic bacteria, are normally present on the human skin. It has also been shown that thorough cleansing of the skin decreases the number of bacteria present. These bacteria introduced into a surgical wound can cause the complication of postoperative wound infection. Surgeons do wear rubber gloves during a surgical procedure, however, it is not infrequent that a surgeon's glove is torn during the course of the operation. This would give bacteria on the skin an opportunity to enter the wound had the surgeon not scrubbed. The arms and elbows are also included in the scrub. These areas are covered by a sterile gown, but usually the sleeves of the gown become wet with blood which increases the ability of bacteria to move from the skin into the operative field. Therefore, the surgeon's hands, elbows and forearms should be scrubbed prior to an operation.
An important complication of surgical procedures is infection. The sources of bacteria causing infection are many. A major source is the surgeon's hands. Even though the surgeon's hands and arms are covered with a sterile gown and gloves, punctures of the gloves by needles or other instruments are frequent. Defects in the gowns are occasionally present. These punctures and defects allow the skin of the surgeon, and thus the bacteria inhabiting the skin, to come in contact with the open wound. Scrubbing the hands and arms of the surgeon before the operation markedly reduces the number of bacteria on his skin and thus the number of bacteria which may come in contact with the wound. The number of bacteria exposed to the wound is a function of the incidence and severity of wound infections. Thus, scrubbing the hands and arms of a surgeon before an operation reduces the danger of wound infections, a serious and frequent surgical complication, and is a good practice.
Early in the nineteenth century Pasteur disclosed his germ theory and linked the microscopic animals with infection. Lister introduced sterilization through the use of carbolic acid in his surgery. By utilizing this method to attempt to sterilize his hands, the patient's skin, and all of the instruments used, his patients had uninfected wounds which healed well and gave good results with decreased mortality and morbidity. Some other surgeons utilized the method with the same results. With the evolution of better and newer cleansing agents and sterilization agents, fewer infections were caused by microorganisms introduced into the wound by the hands of the operator. Therefore, in search of the best method to prevent the surgeons' hands from contaminating the wound, the present method of mechanical cleansing by brushing and chemical sterilization was evolved. In addition to sterile hands, sterile gloves are worn in most countries for added protection of the wound. Because surgical gloves can tear, the hands must be sterile before beginning an operation, and thus this practice is uniform in the United States.
Prior to the last part of the nineteenth century, surgical operations were accompanied by a very high rate of infection and a high mortality. Even before the bacterial etiology of this condition had been firmly established, Semmelweis showed that he could decrease his maternal mortality by two measures. The first was scrupulous care in washing his hands, the second was delivering his patients at home rather than in a hospital. This latter factor was important because of the fact that in a hospital patients met with organisms against which they had not developed an immunity. Since today all operations are done in hospitals where there is unavoidable exposure to a variety of organisms, the surgeon must make sure that he does not carry infection from one patient to another. It has been demonstrated that soap and water is effective in eliminating organisms from the hands for the periods of time involved. This is necessary even though gloves are worn because of the danger of their tearing.
The practice of scrubbing the hands for three minutes with germicidal soap prior to any surgical procedure has become well established. It has been observed that even if this is done, some wound infections occur. It is also true that if the surgeon does not scrub, some patients will not develop wound infections. It can be shown, however, that the number of bacteria isolated per given surface of the surgeon's hand can be reduced by scrubbing in this fashion. The longer the scrub, the fewer the number of organisms isolated. The improvement, per extra minute of scrubbing, diminishes after three minutes, so that the diminishing returns do not justify longer scrubbing. The logic, then, of scrubbing is obvious. The actual benefit could be tested by accumulating data on a large group of similar operations in which alternate cases were done with and without scrubbing. If our premise is true, the incidence of wound infection would be higher in the group operated on without scrubbing.
Rubber gloves prevent hands from contact with patients during surgery. Rubber gloves frequently break. Hands are covered with germs infective to an open wound. Washing hands will wash off some of these germs. Scrubbing hands will get off more of these germs. Scrubbing with an antiseptic is even better. There are a lot of crevices and irregularities to skin. These harbor germs. Therefore scrubbing must be thorough.
The most important reason a surgeon should thoroughly scrub his hands before operating is to guard against infection. Controlled experiments have conclusively shown that a significantly higher incidence of infection occurs following the practice of operating following no scrub as compared to operating after the proper scrub. These tests were performed to evaluate the use of surgical scrub with and without sterile surgical gloves. Four groups of tests were performed: in group one abdominal procedures were performed without gloves and with no scrub; in group two sterile gloves were used with no scrub; in group three procedures were performed following surgical scrub and with no gloves; and in group four both surgical scrub and sterile gloves were used. The highest incidence of infection occurred in group one and the lowest incidence was in group four. There were fewer infections in groups two and three than in group one, but no significant difference was found between these two groups. It was concluded from the study that within the categories studied the most likely method of preventing surgical infection was by the combined use of surgical scrub and sterile gloves.
Infected incisions in the body are harmful. Infections are caused by dirt and germs. Hands not cleaned have germs and dirt on them. Scrubbing the hands removes dirt and germs. Scrubbed hands will not be infective on open wounds. Uninfected open wounds will heal with speed and without morbidity. Patients who have wounds incurred by clean hands will have primary healing. Primary healing is the ideal of all surgical incisions.
Prior to the advent of surgical antisepsis, the postoperative mortality was appalling because of infections. The mortality from infection decreased rapidly following the introduction of clean hands and surgical instruments. By such a simple procedure as washing one's hands before administering to a patient the infection rate dropped suddenly. It is from this basic first principle of antisepsis that the tradition of the surgical scrub evolved. As surgical research and techniques progress, the effectiveness of the surgical scrub is proven again and again. The advent of rubber gloves and antibiotics have still not superceded such a basic procedure. The surgical scrub will continue to be one of the most effective means of preventing wound infection and sepsis.
These are the instructions given to the sample subjects by which the preceding paragraphs were written.

A WRITING EXERCISE

Directions: Today, accepted surgical practice dictates that a surgeon thoroughly scrub his hands before operating. On the blank pages which are attached, you are to organize and write an argument of no less than eight and no more than twelve sentences supporting this practice. Do not try to list all of the reasons you can think of, but take some time to think before writing and try to present as cohesive and logical an argument as you can.
Appendix C

THE EXPERIMENTAL PARAGRAPHS

Paragraph R -- Structure Score = 0.35  
Readability Score = 7.69  
Comprehensibility Index = 304

Infected incisions in the body are harmful. Infections are caused by dirt and germs. Hands not cleaned have germs and dirt on them. Scrubbing the hands removes dirt and germs. Scrubbed hands will not be infective on open wounds. Uninfected open wounds will heal with speed and without morbidity. Patients who have wounds incurred by clean hands will have primary healing. Primary healing is the ideal of all surgical incisions.
Hyperparathyroidism is readily diagnosed in its severe overt form. It is not too difficult to make the diagnosis in a patient with diffuse skeletal deformities and kidneys riddled by stones and calcification. However, this is the exceptional patient, the majority having only a mild to moderate form. In the patient with a mild degree of the disease, the diagnosis is much more difficult and can be made only when it is thought of, and many times only after repeated and prolonged observation and careful study. One patient with bizarre aches and pains came to Walter Bauer's Arthritic Clinic with a normal calcium level when she was first seen. Dr. Bauer, however, suspected that her pains were not truly arthritic, and persisted in his examination in order to arrive at a better diagnosis. Only after repeated observations over a period of three years was he able to reach the tentative diagnosis of hyperparathyroidism. A small adenoma was found and removed. The correction of the hyperparathyroidism in this case has provided one of the most
rewarding results of our series. In general, the number of patients seen at the Massachusetts General Hospital has been due to the painstaking, thorough search and thoughtfulness of our medical colleagues.

In recent years an increasing number of cases of inflammatory fibroid polyps is being recognized. These are not innocent lesions, and it is important for surgeons to be familiar with them. In the past the condition in nearly all the operated patients was mistaken for carcinoma, and hence extensive resections were performed. Fortunately, most of the lesions were located in the distal half of the stomach, where such resections resulted in good functional results. A review of the literature to date reveals no cases of malignancy in these lesions. Two cases are reported herein. In the first the lesion was located on the posterior wall of the fundus, and in the second in the midportion of the stomach on the lesser curvature. Both were large lesions and were diagnosed preoperatively as malignant; both were benign. Bleeding, in the first patient in gradual, unrecognized amounts, and in the second, massive hemorrhage, necessitated surgery.

Nonpenetrating injuries of the abdomen present a challenge to any surgeon. Multiple injuries often mask the abdominal one almost completely. Shock and blood loss are likely to be great and must be treated quickly and effectively. Meanwhile, every method of diagnosis must be used at the moment when it will not harm the patient further. Time and laboratory reports will not settle this question for the surgeon. He must depend on his own hands and brain. The decision as to time of operation when indicated may mean the difference between life and death. Indecision may lose for the patient the only chance he has. The surgeon who meets these cases infrequently must understand the great responsibility he has.

Since the mid-nineteenth century when Semmelweiss and Lister pioneered the concept of antisepsis in and out of the operating room, the practice of scrubbing prior to surgical procedures has evolved through many techniques. The basic tenet is unchanged, however. Dirt, seen or unseen, connotes bacterial contamination. It seems quite obvious, then, that the only way to protect the patient from infection is to thoroughly remove all dirt from anything which may come into direct or indirect contact with the wound. The process of scrubbing is one vital part of obtaining this goal. It is essential to remove all dirt -- i.e., bacteria -- from the hands and forearms. Even though gloves and gowns are worn, the scrubbing assures that torn gloves or wet sleeves will not result in harmful contamination of the wound.
A few more considerations should be mentioned in outlining as I have some of the desirable qualities with which the fledgling surgeon may be well indoctrinated. Another is his willingness to tithe to his profession by donating freely of his time as repayment to society, for the elevation of that profession. He should by this time have been made to realize that society has actually, through its school systems as well as throughout his entire surgical learning experience, contributed the major portion of his education. Throughout his earlier years he may frequently have asked, "Why don't they do this and why don't they do that?" When he is older the obvious answer is, "You are now they." He then realizes completely that surgery is responsible for itself, and that this capability is dependent upon the voluntary dynamic assistance that every mature surgeon must willingly provide. If our new practitioner is sufficiently observing in these concerns he will be awakened to some of the delicacies of the interlocking relation between
self-interest and altruism, and recognize that overt altruism is a disguise for self-interest. Ultimately, examination of his individual conscience will reveal that these qualities are not ambivalent, and that the positive test is the degree of sincerity that exists in altruism. If proved wholesome by this measurement our young surgeon's idols will not be found to have feet of clay, and he will be stimulated to assume his place in the ranks of those striving to serve their profession.

Among ill-chosen words, the dietary unit, the Calorie, is remarkable. Although one thousand times larger than the physicists' calorie which raises 1 ml of water 1° C, its name is exactly the same except that it is spelt with a capital C. I know well that at least half of you had no idea this was so because my questionnaire showed 25 out of 48 fellows of the Royal College of Physicians thought that food and physics calories were identical. It is not your fault -- it is the fault of the word. Until a few years ago I, too, thought they were the same. Then in a thoughtful moment, drinking a large hot cup of tea, I worked out that it was about 10° C hotter than I was and so if my cup held 100 ml I was getting 1,000 calories per cup. In a flash I saw the reason why fat patients never lost weight on my diets and I planned the cold-food cure for corpulence. To find the answer on page 1 of a schoolboys' elementary physics book was humiliating.

---

Bedrest has long been an accepted part of the treatment of gangrene in the lower extremities caused by occlusive arterial disease. This is an apparent contradiction to the increase in circulation in these areas that would accompany walking. Walking was added to the treatment of twenty-two patients with gangrene of the feet or legs resulting from arteriosclerosis, thromboangiitis obliterans, and embolus. This was in addition to normal drug and physical therapies. Walking efforts were repeated hourly with a gradual increase in tolerance. Most patients were eventually able to walk a mile each day, at a pace below that which would cause claudication. Improvement and healing took place in all cases except one. A single patient did not walk as instructed after leaving the hospital and amputation was necessary. This addition of
walking, then, appears to hasten healing and recovery.
It also eases nursing care because the patient is able to
move about and it decreases the length of hospitalization
required.

---

6 Paraphrased from: Foley, William T. "Treatment of
Gangrene of the Feet and Legs by Walking," Circulation, XV
(1957), p. 689.
Almost all of the functions of the body depend upon chemical reactions which need a relatively constant hydrogen ion concentration or pH. To preserve the integrity of these reactions, the pH of arterial blood is normally kept between the very narrow limits of pH 7.40 - 7.43. Small variations of pH can cause physiologic dysfunctions that may add greatly to morbidity. Major changes of pH are incompatible with life unless they are quickly corrected. Changes of pH, often known as acid-base disturbances, are often met among surgical patients. Therefore, it is important that the surgeon have laboratory tools which will give him a correct diagnosis in all cases of acid-base imbalance. Although it has been known for many years that the level of CO₂ in the blood does not give definitive information as to the hydrogen ion concentration, most clinicians still rely principally on changes in blood or plasma CO₂ or CO₂ combining power, for the detection and evaluation of upset acid-base equilibrium. In the past six years, our experience with a simple glass
electrode pH meter has shown that the direct measurement of the arterial blood pH is essential in the proper evaluation and management of acid-base disorders. Dependence on measurement of blood CO\textsubscript{2} alone may lead to serious mistakes in diagnosis. Moreover, an important number of life threatening hydrogen ion imbalances have been met that could only have been detected and measured by the arterial blood pH because in these cases the blood CO\textsubscript{2} value was normal.

---

Appendix D

TEST OF COMPREHENSIBILITY

This test booklet is constructed according to the following plan: The first page contains a medically related paragraph; the second page has questions relating to the information contained in the paragraph. The third page has another paragraph; the fourth has another series of questions, etc.

Your instructions are to (1) note the time, (2) read the first paragraph, (3) record your reading time in seconds in the space at the bottom of the page (reading time is the amount of time it took you to read the paragraph), (4) turn the page and answer the questions WITHOUT REFERRING BACK TO THE PARAGRAPH, and (5) after you finish the questions, turn the page and repeat the procedure. Do this for all nine paragraphs and nine series of questions.

The general rule is: ONCE YOU TURN A PAGE, YOU MAY NOT TURN BACK.
The hardest thing to remember, but it must not be forgotten, is to note the time just before beginning to read and just after finishing reading each paragraph so that you can determine your reading time.

All questions are to be marked true or false depending upon whether or not the statement is an accurate paraphrase of a point made by the author of the paragraph. Write TRUE or FALSE in the space to the left of each question to indicate your decision. Remember, a true statement is an accurate restatement of a point made in the paragraph; a false statement may be either an inaccurate restatement or a statement you know to be true from experience that was not made in the paragraph by its author.

If you do not thoroughly understand what you are to do, ask questions before beginning. Any deviation from the outlined procedure will completely invalidate your paper.
Questions on Paragraph -- R

1) Infection is harmful.
2) Dirt causes infection.
3) Scrubbing kills germs.
4) Scrubbed hands are not infective.
5) Primary healing is to be preferred.
6) Wounds incurred by unscrubbed hands result in secondary healing.
7) Uninfected wounds heal less quickly.
8) Morbidity is often result of infection.
9) Dirt carries bacteria.
10) Primary healing results from clean hands.
The more severe the hyperparathyroidism, the easier the diagnosis.

Most cases of hyperparathyroidism are severe.

Normal calcium level always contra-indicates hyperparathyroidism.

The reported case was a result of an adenoma.

The usual case of hyperparathyroidism is fairly easily diagnosed.

Skeletal deformities and kidney stones are symptomatic of hyperparathyroidism.

Once hyperparathyroidism is suspected, final diagnosis is rather quick.

The diagnosis of the reported case took three months.

Hyperparathyroidism is easily corrected once it is diagnosed.

Hyperparathyroidism is not a common adult disease.
Questions on Paragraph -- U

1) Inflammatory fibroid polyps are often diagnosed as carcinomas.

2) Most of the lesions mentioned were on the proximal half of the stomach.

3) The discussed lesions were often malignant.

4) The two cases mentioned were of small, benign lesions.

5) Bleeding was, in both cases, the immediate reason for surgery.

6) Inflammatory fibroid polyps are generally innocent lesions.

7) Because these lesions are mistaken as carcinoma, resections are usually performed.

8) Both reported cases were diagnosed preoperatively as malignant.

9) Both reported cases had massive hemorrhage.

10) Resections, when performed, do not give functional results.
Questions on Paragraph -- D

1) Indecision contributed to mortality.
2) Time of operating is a vital decision.
3) Shock and blood loss are likely to be great.
4) Other injuries often hide an abdominal one.
5) Nonpenetrating injuries of the abdomen are difficult to diagnosis.
6) In these cases the surgeon assumes great responsibility.
7) Laboratory reports will not settle the surgeon's questions.
8) Treatment of shock and blood loss is first priority treatment.
9) Each diagnostic procedure should be used when it will not harm the patient further.
10) Laboratory reports are of no value.
11) These injuries are met infrequently.
12) Penetrating abdominal injuries are more common than nonpenetrating ones.
13) The surgeon must not rely on his own ideas but on lab reports.
14) All possible diagnostic methods should be employed immediately.
15) Nonpenetrating abdominal injuries are never immediately obvious.
Questions on Paragraph -- N

1) Semmelweiss was a surgeon.

2) Lister pioneered the idea of antisepsis.

3) Scrubbing removes bacteria.

4) Dirt carries bacteria.

5) Gloves and gowns are the only necessary protection against infection.

6) The surgical scrub is the same today as it was a century ago.

7) Dirt may not be visible.

8) Scrubbing is an important, but not a vital, part of the prevention of wound infection.

9) Scrubbing kills bacteria.

10) Sterile gloves and gowns are not protection enough.
Questions on Paragraph -- 0

1) Society provides small contribution to a surgeon's education.

2) Progress of surgery depends upon the efforts of surgeons.

3) Self-interest and altruism are unrelated.

4) Surgeons must donate time and effort to the advancement of their field.

5) The older surgeons are more responsible for the field than younger ones.

6) Self-interest is a disguise for overt altruism.

7) Altruism is measured by sincerity.

8) Society contributes a fiscally minor portion of the cost of a surgeon's education.

9) Surgery is independent unto itself.

10) The young surgeon must tithe to his profession to repay it for his opportunities.
Questions on Paragraph -- E

1) The physicist's calorie is 1,000 times larger than the dietary unit of the same name.

2) The physicist's calorie is spelled with a capital C.

3) About 3/4 of the questioned physicians thought the two units were not identical.

4) The physicist's calorie raises 1 ml of water 1°C.

5) The dietary calorie raises 1 ml of water 1°C.

6) The calorie is a measure of force.

7) The physicist's unit and the dietary unit are interchangeable.

8) Cold food contains more calories than hot food.

9) Dieting can cure corpulence.

10) The author was surprised to find his colleagues uninformed about these units.
Questions on Paragraph -- A

1) Treatment of gangrene of lower extremities usually includes bedrest.

2) Walking increases circulation in the legs.

3) The reported treatment was used with over 100 patients.

4) In all cases the gangrene resulted from occlusive arterial disease.

5) The limit on walking speed was claudication.

6) All cases showed improvement and no amputations were performed.

7) All cases of lower extremity gangrene result in amputation if bedrest is used in the treatment.

8) Each patient walked a mile each day.

9) The reported treatment eases nursing care.

10) The reported treatment reduces the number of necessary amputations.
Questions on Paragraph -- X

1) Hydrogen ion concentration is expressed by pH.

2) Hydroxyl ion concentration is expressed by pOH.

3) Neutrality is when pH = 7.00.

4) Blood pH is normally slightly basic.

5) Blood pH changes are common in surgical patients.

6) CO₂ level is a good measure of blood pH.

7) CO₂ level is a common measure of blood pH.

8) An acid-base imbalance is a pH change in arterial blood.

9) Minor changes of pH are incompatible with life.

10) Even small pH changes add to morbidity.

11) The author suggests a titrometer for measuring blood pH.
Appendix E

TECHNICAL TERMS USED IN THE EXPERIMENTAL PARAGRAPHS

<table>
<thead>
<tr>
<th>Term</th>
<th>Uses</th>
<th>Term</th>
<th>Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>abdomen</td>
<td>1</td>
<td>CO₂</td>
<td>5</td>
</tr>
<tr>
<td>abdominal</td>
<td>1</td>
<td>contamination</td>
<td>1</td>
</tr>
<tr>
<td>acid-base</td>
<td>4</td>
<td>corpulence</td>
<td>1</td>
</tr>
<tr>
<td>adenoma</td>
<td>1</td>
<td>1° (degree)</td>
<td>2</td>
</tr>
<tr>
<td>amputation</td>
<td>1</td>
<td>diagnosed</td>
<td>1</td>
</tr>
<tr>
<td>antisepsis</td>
<td>1</td>
<td>diagnosis</td>
<td>7</td>
</tr>
<tr>
<td>arterial</td>
<td>4</td>
<td>dietary</td>
<td>1</td>
</tr>
<tr>
<td>arteriosclerosis</td>
<td>1</td>
<td>distal</td>
<td>1</td>
</tr>
<tr>
<td>arthritic</td>
<td>2</td>
<td>dysfunctions</td>
<td>1</td>
</tr>
<tr>
<td>bacterial</td>
<td>2</td>
<td>electrode</td>
<td>1</td>
</tr>
<tr>
<td>benign</td>
<td>1</td>
<td>embolus</td>
<td>1</td>
</tr>
<tr>
<td>C (centigrade)</td>
<td>2</td>
<td>fibroid</td>
<td>1</td>
</tr>
<tr>
<td>calcification</td>
<td>1</td>
<td>fundus</td>
<td>1</td>
</tr>
<tr>
<td>calorie</td>
<td>4</td>
<td>gangrene</td>
<td>2</td>
</tr>
<tr>
<td>carcinoma</td>
<td>1</td>
<td>germs</td>
<td>2</td>
</tr>
<tr>
<td>claudication</td>
<td>1</td>
<td>hemorrhage</td>
<td>1</td>
</tr>
<tr>
<td>clinicians</td>
<td>1</td>
<td>hyperparathyroidism</td>
<td>3</td>
</tr>
</tbody>
</table>

163
<table>
<thead>
<tr>
<th>Term:</th>
<th>Uses:</th>
<th>Term:</th>
<th>Uses:</th>
</tr>
</thead>
<tbody>
<tr>
<td>incisions</td>
<td>2</td>
<td>physiologic</td>
<td>1</td>
</tr>
<tr>
<td>infected</td>
<td>1</td>
<td>plasma</td>
<td>1</td>
</tr>
<tr>
<td>infection</td>
<td>2</td>
<td>polyps</td>
<td>1</td>
</tr>
<tr>
<td>infective</td>
<td>1</td>
<td>practitioner</td>
<td>1</td>
</tr>
<tr>
<td>inflammatory</td>
<td>1</td>
<td>primary healing</td>
<td>1</td>
</tr>
<tr>
<td>ion</td>
<td>3</td>
<td>resections</td>
<td>2</td>
</tr>
<tr>
<td>lesser curvature</td>
<td>1</td>
<td>shock</td>
<td>1</td>
</tr>
<tr>
<td>lesions</td>
<td>5</td>
<td>skeletal</td>
<td>1</td>
</tr>
<tr>
<td>malignancy</td>
<td>2</td>
<td>surgeon</td>
<td>9</td>
</tr>
<tr>
<td>ml (milliliter)</td>
<td>2</td>
<td>surgery</td>
<td>2</td>
</tr>
<tr>
<td>morbidity</td>
<td>2</td>
<td>surgical</td>
<td>5</td>
</tr>
<tr>
<td>nonpenetrating</td>
<td>1</td>
<td>therapies</td>
<td>1</td>
</tr>
<tr>
<td>occlusive</td>
<td>1</td>
<td>thromboangiitis obliterator</td>
<td>1</td>
</tr>
<tr>
<td>pH</td>
<td>8</td>
<td>uninfected</td>
<td>1</td>
</tr>
<tr>
<td>physicians</td>
<td>1</td>
<td>wound</td>
<td>4</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


Buswell, Guy T. "Relationship between Rate of Thinking and Rate of Reading," School Review, LIX (September, 1951), pp. 339-46.

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


_________. A Formula for Predicting Readability. Columbus, Ohio: Bureau of Educational Research, The Ohio State University, 1948.

165


Heidbreder, E. "An Experimental Study of Thinking," *Archives of Psychology*, XI, Number 73, (1924).


Maier, N. R. F. "Reasoning in Humans I: On Direction," 
*Journal of Comparative Psychology,* X (1930), 
pp. 115-43.

———. "Reasoning in Humans II: The Solution of 
Problem and Its Appearance in Consciousness," 
*Journal of Comparative Psychology,* XII (1931), 
pp. 181-94.

———. "Reasoning in Humans III: The Mechanisms 
of Equivalent Stimuli and of Reasoning," *Journal 
of Experimental Psychology,* XXXV (October, 1945), 
pp. 349-60.

Michaels, H.S. "Logical Approach to Composition," 
*College English,* XII (April, 1951), pp. 390-93.

Miller, G. A. "Speech and Language," Chapter 21, 
pp. 789-810 in: Stevens, S. S. *Handbook of Experi­

Minton, A. "Thinking-Composition," *English Journal,* XL 
(January, 1951), pp. 7-11.

Moore, Robert H. *Effective Writing.* New York: 
________. The Morgan Test of Logical Reasoning.  


I, Richard Homer Barbe, was born on December 23, 1933. I received my elementary and secondary-school education in Wellington, Ohio. I entered Miami University in 1951 with an N.R.O.T.C. scholarship, and in 1955 I received a B.S. in Education degree and a commission in the U.S. Navy. During my senior year at Miami I served as a part-time instructor of the physical sciences in the College of Education. I served in the navy until 1958, finishing my tour of duty as a destroyer's engineering officer. I entered the Graduate School of The Ohio State University in 1958 and was a full-time student until the fall of 1961. In 1959 I received a Master of Arts degree. While at Ohio State I was a research assistant to Dr. Edgar Dale from 1958 to 1960 and a research assistant to Prof. George Kienzle from 1960 to 1961. I am now employed as an assistant professor and research associate in the School of Journalism of The Ohio State University.