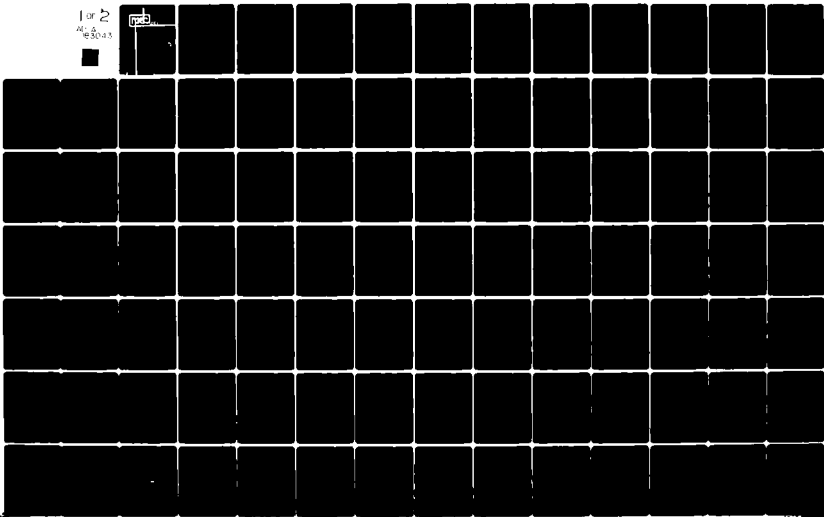
AD-A083 043 NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER SAN D--ETC F/G 5/9
TRI-SERVICE LITERACY AND READABILITY: WORKSHOP PROCEEDINGS.(U)
MAR 80 T E CURRAN
UNCLASSIFIED NPRDC-SR-80-12 NL

1 OF 2
AD A
083043



npdc LEVEL II

NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER

NPRDC SR 80-12 MARCH 1980

TRI-SERVICE LITERACY AND READABILITY: WORKSHOP PROCEEDINGS

NPRDC Special Report 80-12 (12) March 1980

TRI-SERVICE LITERACY AND READABILITY: WORKSHOP PROCEEDINGS

Edited by
Thomas E. Curran

Reviewed by
R. E. Blanchard

Released by
Donald F. Parker
Commanding Officer

Navy Personnel Research and Development Center
San Diego, California 92152

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NPRDC-SR-80-12	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) TRI-SERVICE LITERACY AND READABILITY: WORKSHOP PROCEEDINGS		5. TYPE OF REPORT & PERIOD COVERED Final Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Thomas E. Curran (Editor)		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ZF55 522 003
11. CONTROLLING OFFICE NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		12. REPORT DATE Mar[redacted]80
		13. NUMBER OF PAGES 100
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Readability
Literacy
Comprehensibility
Motivation
Reading Grade Level (RGL)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Tri-Service Literacy and Readability Workshop was held to encourage closer coordination among persons working in the field of readability and literacy. The workshop was funded by the Naval Technical Information Presentation Program through its project at NAVPERSRANDCEN. It was organized and coordinated by the NAVPERSRANDCEN project officer and hosted by the Air Force Human Resources Laboratory (AFHRL) at Lowry Air Force Base, Denver.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

390772

FOREWORD

The Tri-Service Literacy and Readability Workshop was held to encourage closer coodination among persons working in the field of readability and literacy. The workshop was funded by the Naval Technical Information Presentation Program through its project at NAVPERSRANDCEN. It was organized and coordinated by the NAVPERSRANDCEN project officer and hosted by the Air Force Human Resources Laboratory (AFHRL) at Lowry Air Force Base, Denver.

Every attempt was made to invite representatives of all services and persons involved in all aspects of the readability-literacy field. It is believed that the most important outcome of the workshop was the potential for closer communication ties among the services and the promise of a much more coordinated effort in the future. The recommendations presented in the summary and in the "Recapitulation" paper (p. 86) are directed toward technical data developers, especially those working for or with the uniformed services.

The papers presented in this document essentially are verbatim reproductions of those presented at the workshop.

DONALD F. PARKER
Commanding Officer

CONTENTS

INTRODUCTION

Problem

American society in general, and the Armed Forces in particular, are increasingly concerned with the declining skills of young people coupled with the increasing demands being placed upon them. The problem has two general facets: "literacy" (a person's ability, particularly in regard to reading) and "readability" (the difficulty of the materials to be read). In society in general, this problem is evidenced by the decreasing scores students obtain on the Scholastic Aptitude Test (SAT), the widely used college entrance examination. In the military, the problem is evidenced by the increasing number of personnel who attrite because they cannot read required materials.

Purpose

Persons concerned with alleviating the literacy/readability problem in the military felt that duplication of effort might well be occurring and that current methods of communicating information were not adequate. Therefore, as a first step toward opening lines of communication and fostering coordination and cooperation, it was decided to hold a 3-day Tri-service Literacy and Readability Workshop at Lowry Air Force Base, Denver, CO in August 1978. The purpose of this workshop was to allow persons engaged in research and/or application in the literacy/readability field to discuss mutual problems and potential solutions, and to apprise one another of specific efforts being undertaken.

Attendees

Twenty-two persons participated in the workshop. The majority of the participants are affiliated directly with the Armed Services in one capacity or another. Of these, approximately half were uniformed service personnel. The remainder were persons from private industry, who have been or are conducting work on literacy/readability under government contract.

Format of the Workshop

Doctor George R. Klare of Ohio University, a prominent expert in the field of readability, delivered a keynote speech and acted as a general advisor to the group. The remainder of the workshop was organized around a set of working papers prepared and delivered by persons involved in both literacy and readability and in both research and application. Each paper was followed by a general discussion. On the final afternoon, a discussion was held to reach consensus on the many ideas arising during the workshop.

Conclusions and Recommendations

The conclusions and recommendations that resulted from the workshop are listed below and are discussed in a recapitulation of the workshop written by Dr. Klare (pp. 86-89).

1. Every attempt should be made to clarify the use of the term "Reading Grade Level (RGL) to eliminate the misunderstanding that currently exists.

2. Attempts to develop new readability formulas should be resisted unless new and better index variables in written materials can be identified.

3. Caution should be used in the application of readability formulas; if the formula was developed with a relatively naive norm group, it may be very misleading to apply it to materials intended to be read by experienced personnel.

4. Efforts to determine the effect of "technical terminology" on readability formulas and on the nature of a man's "technical reading ability" as he gains experience in his field should be continued.

5. Efforts to develop a cost-effective computerized text authoring/editing system should be continued. Also, they should be closely coordinated among the three services, in view of the potentially significant cost savings that might be realized by eliminating overlapping efforts.

6. Increased use of "performance criteria" rather than "verbal comprehension criteria" should be made in the study of literacy and readability problems. Associated with this, an attempt should be made to develop methods for "unobtrusive measurement" in the field.

7. Attempts to develop readable, usable, and effective writer's guides should be continued. Such guides, if available, would reduce the chances of writers making mechanical changes in their work that may satisfy a readability formula score but in fact do not improve the comprehensibility of the material.

8. Efforts to obtain a better understanding of the concept and the characteristics of "comprehension" should be continued.

A POSSIBLE FRAMEWORK FOR THE STUDY OF READABILITY

George R. Klare
Department of Psychology, Ohio University
Athens, OH

ABSTRACT

This paper presents the chief features of a possible framework for the study of readability. A brief historical introduction precedes a look at the relationship of readability values to reader comprehension scores, using the framework.

The first section, "Correlation Need Not Imply Causation," contrasts readability research of two major types: (1) prediction, or use of readability formulas to predict comprehensibility, and (2) production, or the attempt to write comprehensibly.

The second section answers the question of the relationship of readability to comprehension as "Sometimes Yes, Sometimes No." The major reasons for this answer often lie in the handling of the research situation, as several studies show.

The third section indicates a need to go "Back to the Drawing Board" for a clearer understanding of the concept of comprehension.

Introduction

First, I would like to present briefly the chief features of the framework for the study of readability. Also, I would like to review the relationship of readability values to the comprehension levels obtained with readers. In the process, I would like to examine some research results—some of mine, and, hopefully, some of yours--in the discussion that follows. I will try to illustrate three points concerning the relation of readability to comprehension: (1) Correlation Need Not Imply Causation; (2) Sometimes Yes, Sometimes No; and (3) Back to the Drawing Board.

Now, let me lay the foundation for the possible framework for studying readability. First, let me emphasize the word "possible." Your comments and suggestions can help me in modifying this beginning framework where needed. Second, let me apologize if I repeat parts of the framework already familiar to you. I will try to be brief and yet be complete enough to make sense.

As a consequence, I will present only the chief features of the framework with little supporting rationale. For one thing, the rationale appears more fully in a recent paper called "A Second Look at the Validity of Readability Formulas" (Klare, 1976). For another, I want to save enough time, as indicated earlier, to look at some studies from the point of view of the framework.

Most readability research (as opposed to application) seems to me to have one of two major purposes:

1. Distinguishing between samples of writing as likely to be more versus less readable to readers. I have been referring to this as the "prediction" of readable writing, as exemplified by the development of readability formulas. Figure 1 presents the best-known formula, Rudolf Flesch's "Reading Ease" (1948), along with the revision for Navy enlisted personnel derived by Kincaid, Fishburne, Rogers, and Chissom (1975).

$$RE = 206.835 - .846\, wl - 1.015\, sl$$

where: RE = Reading Ease
wl = syllables per 100 words
sl = words per sentence

Revision

$$GL = .39\,(\text{words/sentences}) + 11.8\,(\text{syllables/word}) - 15.59$$

where: GL = grade level

Figure 1. The Flesch Reading Ease Formula and the revision by Kincaid, Fishburne, Rogers, and Chissom.

2. Deciding how to write readably, or change writing to make it more readable. I have been referring to this as the "production" of readable writing, as exemplified by the development of manuals for readable writing. Examples are Kern, Sticht, Welty, and Hauke's "Guidebook for the Development of Army Training Literature" (1975) and my "A Manual for Readable Writing (1975)."

Some basic differences appear between prediction and production when it comes to doing research, particularly validity research. I have summarized them in a 2 x 2 table that was first presented in the "Second Look" paper (Klare, 1976) and is pictured in Figure 2.

	Prediction of Readable Writing	Production of Readable Writing
Readability Variables	Index	Causal
Validity Check	Correlational	Experimental

Figure 2. Two approaches to research on the validity of readability measures.

With that brief introduction, we can turn to the first point concerning the relationship of readability values to reader comprehension scores.

Correlation Need Not Imply Causation

The common goal of prediction research is to discover language variables that correlate highly with comprehension scores. These need only be index variables. They could have a causal relationship, but they need not have, as long as they serve as efficient indices. If, in other words, they are relatively simple and can be counted easily and reliably by hand or by computer.

Prediction research has, in general, been quite successful. Simple index variables in language have been found to correlate highly with complex, probably causal variables, as the following examples show.

1. Number of morphemes per 100 words of text appears to be a cause of semantic difficulty. Yet the much simpler index of syllables per 100 words correlates .95 with this count (Coleman, 1971).

2. The sum of Yngve word depths per sentence has been considered a cause of syntactic difficulty. Yet the number of words per sentence correlates .99 with this count (Bormuth, 1966), and is much simpler to use.

3. Number of propositions has been considered a cause of conceptual difficulty (particularly as this relates to memory for meaning) (Kintsch, 1974). Yet the number of syllables yields a higher correlation with passage complexity ratings--.70--than does number of prepositions--.60--(Kling & Pratt, 1977).

The readability formulas based upon these simple index variables also show high correlations with comprehension criteria. Let me use several of Coleman's formulas (1971) as examples.

1. Coleman's two-variable formula uses percentage of one-syllable words and number of sentences in 100 words as predictors. This formula yielded a correlation of .898 with cloze percentage correct.

2. This value held up extremely well in cross-validation, yielding a value of .88 (Szalay, 1965).

Furthermore, when Coleman added variables to those above, the predictive power changed very little.

3. Adding number of pronouns raised the correlation to only .903, and this went up to .910 when number of prepositions was also added.

4. The cross-validation values, once again, reflected this lack of change, being .87 and .89 respectively.

What are the implications of the prediction research? First, the good news. Prediction of comprehension scores with simple readability variables can yield very high correlations--certainly much higher than most kinds of psychological or educational prediction, such as success on the job or school (or college) grade achievement. Consequently, little reason other than academic justifies further protracted search for better index variables, at least as far as the typical prediction task is concerned.

Second, the bad news. Adding new and different variables, however promising they may sound, usually turns out to be a frustrating task since correlations with comprehension scores seldom go up very much. Certainly they go up very little compared to the extra work entailed in using the added variables. Only such special needs as research or cross-checking justify the extra effort.

Third, there is more bad news.

First, correlation coefficients are, in some ways, tricky statistical measures. The magnitude of a coefficient depends too much upon the range of difficulty in the criterion and the range of ability in the subjects involved. "Ability," as used here, should include such characteristics as interest in, and prior knowledge of, the content in the criterion passages. Consequently, the high correlations in prediction research are, at least to some extent, a function of the wide variety of the passages used and/or the wide range of the subjects used. This should warn us that correlation may not signal causation in the real world.

Second, comprehension is at present too poorly understood to be measured very satisfactorily--at least as far as agreement is concerned. For example, correlations with cloze scores tend to be significantly higher than with multiple-choice scores. But, is cloze really a better measure of comprehension? Again, this indicates that correlation and causation may be quite different in fact.

I'll return to the matter of comprehension later since I feel it a very serious one. For the moment, though, let me illustrate that "correlation need not imply causation" by presenting a summary of 36 <u>production</u> studies in Table 1. These studies, all of them, were designed to relate changes in readability values to changes in comprehension scores. All were experimental in nature.

Note that in this summary table:

1. "Positive" really means that there <u>was</u> a statistically significant relationship.

2. "Negative" really means that the relationship <u>was not</u> statistically significant.

3. "Mixed" indicates that some analyses showed significant relationships and some did not.

High correlations in prediction studies, then, do not necessarily imply causative relationships in production studies.

Why not?

In the attempt to answer this question, I examined the 36 studies in great detail. I looked first at 40 characteristics in each, then cut this to 28 when I found that I could not get information on all 40. The 28, in turn, could be categorized under 5 major headings, and these form the basis for the possible framework I would like to suggest for the study of readability. It is this framework, or model, that I believe helps to tell why the question of the relationship of readability values to comprehension scores must be answered.

Table 1

Summary of Relationships of 36 Experimental Studies of the Effect of Readability Variables Upon Comprehension Scores

Relationship	All Studies	Published Studies	Theses or Dissertations
Positive	19	6	13
Mixed	6	3	3
Negative	11	0	11
	36	9	27

Sometimes Yes, Sometimes No

The framework presented in Figure 3 comes from the "Second Look" paper mentioned earlier (Klare, 1976). In the attempt to evaluate this model, several of us at Ohio University have been, jointly or singly, testing predictions from it. I have also worked with Tom Curran in using it to try to explain existing readability studies, and more recently with Tom Duffy in connection with a study of his. I am interested in refining the model as well as in explaining results, particularly those that seem contradictory, in the literature.

Time will not permit me to go into the contents of all of the boxes in the framework, or to present the details of the studies here. Let me, therefore, invite your questions later or refer you to the "Second Look" paper itself. One more preliminary--concerning the term "Interacting with" which you see spread liberally throughout the model. They have been, whatever their appearance, put there to fill a genuine need, as I'll hope to show.

As a first illustration of the interacting nature of the other categories with readability values, let me cite a study by Warren Fass and Gary Schumacher (in press). A little background should put it in perspective.

In the "Second Look" paper, I suggested the likelihood that readability changes would significantly affect comprehension scores; that is, they would be lowered when two classes of motivation-related conditions were present in the test situation: One or more conditions that raised the level of motivation, such as promise of reward, threat of punishment, or even the experimental (test) situation itself, combined with one or more conditions that allowed the increased motivation to have an effect upon behavior, such as liberal reading and/or testing time.

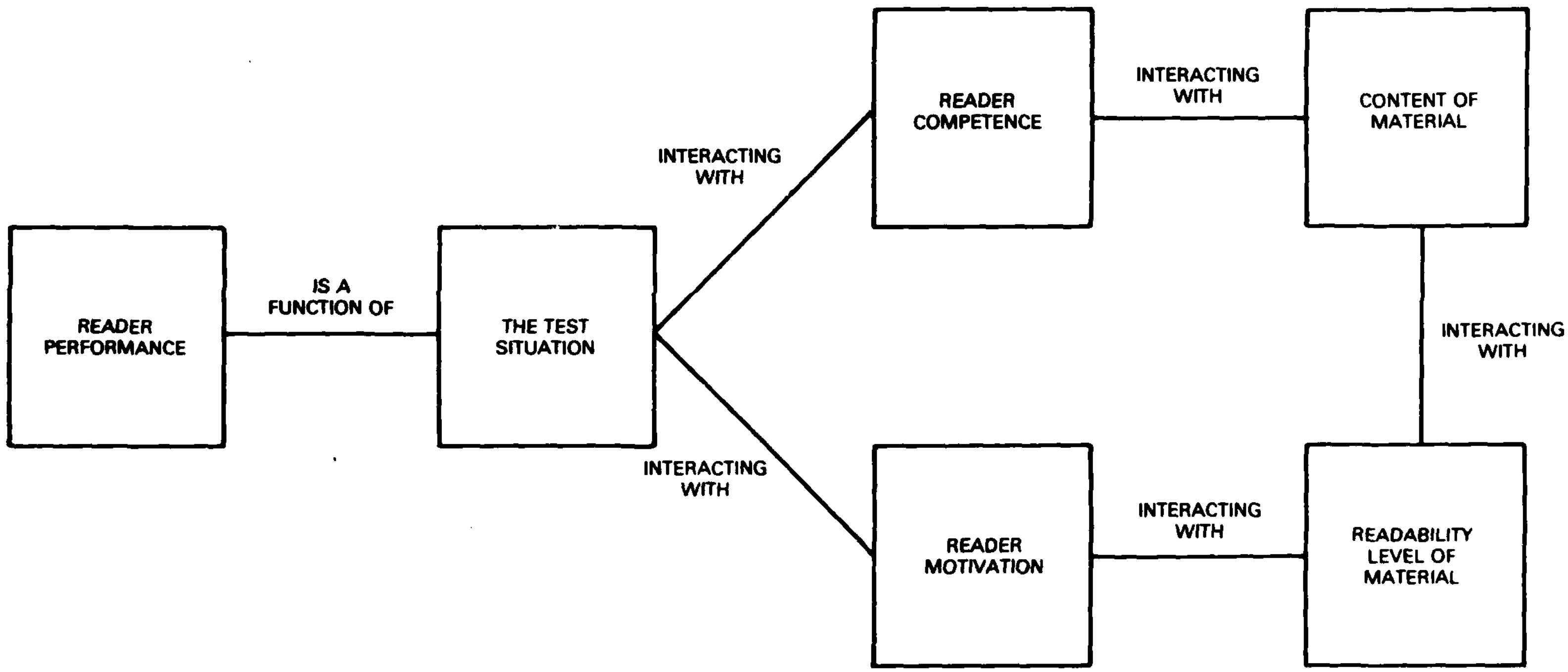


Figure 3. Some major factors interacting with readability measures in validity studies.

Figure 4 provides an expansion of "The Test Situation" box in the model.

A. Motivating Factor(s) + B. Effectiveness Factor(s)

A. Motivating Factor(s)	B. Effectiveness Factor(s)
1. Type of instructions	1. Liberal reading time
2. Threat present	2. Liberal testing time
3. Experimenter uses own class and classroom	3. Text present during testing
4. Payment for participation	4. Opportunity for rereadings

C. Level of readability of test (in relation to readability of text)

1. Multiple-choice measure
2. Cloze comprehension measure

Figure 4. Some test situation considerations.

With that background in mind, let me describe the Fass-Schumacher study very briefly.

1. Two versons of a 1000 ± word passage on enzymes were used, one at a Reading Ease score of 61 (around 8-9 grade level) and one at a score of 31 (around college level).

2. Eighty college freshmen read the more readable version for 10 minutes (a liberal reading time) and another 80 read the less readable version for 10 minutes. They were then, in both cases, given 15 minutes to complete a 15-item multiple-choice comprehension test (again a liberal time). This provided the "readability" effect in a factorial design.

3. Eighty of the 160 freshmen were told the five highest scores would get five dollars each. The other 80 subjects were unaware of any pay-off (they did get some points toward their course grades, but unfortunately this has proved to be a rather ineffective motivator in the past). This provided the "motivation" effect in the factorial design.

4. Eighty of the subjects were told to underline key words while reading; the other 80 were told to make no marks while reading. This provided the "task" effect in the factorial design.

Table 2 provides means for the three major variables; readability, motivation, and task. Table 3 shows that each of the variables produced significant results, but it also shows the interactive effects of motivation. The a priori contrasts show that the scores on the more readable (easier) version were significantly higher than those on the less readable (harder) version under conditions of lower motivation. Under conditions of higher motivation, however, the difference is no longer significant. This is the interaction effect predicted by the model.

Table 2

Adjusted Mean Number of Questions Answered Correctly as a Function of Motivation, Readability, and Task Manipulation

Motivation	Readability Level Easy	Hard
Higher:		
Underline	9.90	9.15
Read Only	8.14	6.45
Lower:		
Underline	8.56	6.77
Read Only	7.93	5.73

Table 3

Analyses of Motivation, Readability, and Task Scores

Analysis of covariance

Motivation, $F(1,151) = 4.25$, $p < .05$
Readability, $F(1,151) = 9.01$, $p < .01$
Task Scores, $F(1,151) = 8.20$, $p < .01$

Planned a priori contrasts (Dunn's Procedure)

Easy vs. Hard, Lower Motivation, $F(1,151) = 9.15$, $p < .01$
Easy vs. Hard, Higher Motivation, $F(1,151)$ = nonsignificant

I would like to describe one more study, one I made with Eileen Entin (Entin & Klare, submitted for publication). Once again, let me provide a bit of background. To begin, let me illustrate "cloze procedure," a term I have used several times earlier. Figure 5 illustrates both "standard" cloze procedures (with uniform blanks replacing the missing words) and the experimental "dash" version (with underlines representing the letters of the missing words).

The only banking __________ in which a guaranty- __________ provision is actually incorporated __________ the present time is __________ of Canada. Acccording to __________ terms of the banking __________ of 1890, the notes __________ the bank are made __________ first charge upon all __________ assets of the issuing __________; also each stockholder may __________ forced to contribute his __________ and a like amount __________ cash.

The only banking _ _ _ _ _ _ in which a guaranty- _ _ _ _ provision is actually incorporated _ _ the present time is _ _ _ _ of Canada. According to _ _ _ terms of the banking _ _ _ of 1890, the notes _ _ the bank are made _ first charge upon all _ _ _ assets of the issuing _ _ _ _; also each stockholder may _ _ forced to contribute his _ _ _ _ _ _ and a like amount _ _ cash.

Answers: system, fund, at, that, the act, of, a, the, bank, be, shares, of

Figure 5. Standard and experimental versions of cloze tests.

Now, let me describe the study briefly.

1. We began by comparing the standard version of cloze procedure with the experimental dash version in terms of correlation of both with multiple-choice comprehension scores and with readability values.

2. The multiple-choice comprehension items were those provided with Nelson-Denny Reading Test passages. The readability values were RE scores on seven Nelson-Denny passages.

3. We used several control groups to be sure, for example, that taking a cloze test on a passage would not affect a multiple-choice test score taken 1 week later. It did not. Since there were no unintended effects, we can forget about the control groups for the moment to reduce complexity and save time. We can also forget the differences between the two cloze groups, since the effects were similar with both.

4. All that we need to note in Table 4 is that:

- CL_1 and CL_2 values are based upon correlations using mean scores on two groups tested with the cloze procedure on seven Nelson-Denny passages.

- RE values are based upon correlations using the Flesch Reading Ease formula on the same seven passages.

- MC_1 and MC_2 values are based upon correlations for two groups who took the multiple-choice tests after reading the seven Nelson-Denny passages. The label "Uncorr" signifies that the scores were used just as they came from testing (i.e., they were "Uncorrected").

- Each of the groups contained from 30 to 60 freshmen.

The results of the study appear in Table 4. Note that, across the top of the table, the correlations between RE values and cloze scores are quite high--especially when you consider that the N here was only seven (i.e., seven passages). Also note, however, that the correlations between the RE values and multiple-choice scores are essentially zero (if not worse).

Table 4

Correlation Coefficients, Based Upon Seven Nelson-Denny Comprehension Test Passages, Between Flesch Readability (RE), Cloze (CL), and Multiple-Choice Scores Uncorrected (MC-Uncorr) and Corrected (MC-Corr) for Prior Knowledge

Item	RE	CL_1	CL_2	MC_1-Uncorr	MC_2-Uncorr
RE		.68	.74	-.11	.01
CL_1			.97	-.22	-.08
CL_2				-.17	-.04
MC_1-Corr	.41	.48	.45	.34	.49
MC_2-Corr	.44	.45	.50	.29	.45

This was puzzling, at least for a time. But Figure 6, which is an expansion of the "Content of Material" box in the model, provided a possible answer. As this figure suggests, the effect of readability should be reduced to the extent that the text used presents relatively little new information.

A. New information (in relation to reader knowledge).

B. Interest-value (in relation to reader interests).

C. Nature of content (in relation to reader intellectual level).

D. Maturity of content (in relation to reader maturity).

Figure 6. Some content considerations.

To test this hypothesis, we gathered a new group of 50 freshmen and gave them the Nelson-Denny multiple-choice questions before they read the Nelson-Denny passages. This gave us a "prior knowledge" score for each question, which we then subtracted from the scores of the subjects who had taken the questions after reading the passages.

Now look at the correlations based upon these "corrected" multiple-choice scores. The correlations have gone up to a respectable size, especially considering again that they are based on only seven cases--seven passages.

Of course, it would have been better had we been able to obtain "prior knowledge" scores on the same subjects rather than a new group; that is, the corrections would have been better. But, of course, we could not do that in this instance. I hope someone will, however, because when two measures of comprehension give such different results, it is a cause for concern beyond just that of testing the validity of readability measures. And it is especially serious when items give better before scores than after scores, which happened with several items. And, of course, when it occurs on a test as long-used and as widely used as the Nelson-Denny Reading Test.

That is why I say, when it comes to comprehension, we need to go back to the drawing board.

Back to the Drawing Board

No matter how careful the work in relating readability to comprehension, our lack of understanding of reading comprehension will always place a limit on our progress.

The best way I know to show you quickly the confusion in this area is to give you a set of examples. I began collecting definitions of "comprehension" and "understanding" several years ago and would like to share some of them with you.

I decided that I would need two stones to kill this one bird, so I have a set of historically-oriented definitions, by famous definers; and recent definitions, by competent but less-well-known definers.

They are set up as two matching tests designed to make my point. They should be of additional interest to persons with a psychological background, as you'll see.

Let me ask you to look at Matching Test 1 (Figure 7) first and pair the definitions above with the names of definers below.

Now take a few more minutes and do the same for Matching Test 2 (Figure 8).

When you have finished, turn to page 16 to check your answers.

I hope you noticed, first of all, the wide disparity in the definitions. If it were not for the unavoidable clues, you would probably have had trouble deciding what was being defined. I hope also that those of you with a psychological background did significantly better on Matching Test 1 than on Matching Test 2. Why?

Please read the five definitions of comprehension (understanding), and decide which of the five authors listed below them is being quoted in each.

1. The listener can be said to understand a speaker if he simply behaves in an appropriate fashion . . . . In "instruction," we shall see that he understands to the extent that his future behavior shows an appropriate change. These are all ways in which we are said to "understand a language"; we respond according to previous exposure to certain contingencies in a verbal environment.

2. Those who have read of everything are thought to understand everything too, but it is not always so. Reading furnishes the mind only with materials of knowledge, it is thinking that makes what we read ours.

3. That the general meaning dawns upon the reader precedent to the full sentence-utterance is evidenced by the many cases in which variant words of equivalent meaning are read, and also by the comparative ease with which a reader may paraphrase the thought of what he reads . . . . It is of the greatest service to the reader or listener that at each moment a considerable amount of what is being read should hang suspended in the primary memory of the inner speech. It is doubtless true that without something of this there could be no comprehension of speech at all.

4. These deep structures, along with the transformation rules that relate them to surface structure and the rules relating deep and surface structures to representations of sound and meaning, are the rules that have been mastered by the person who has learned a language. They constitute his knowledge of the language; they are put to use when he speaks and understands.

5. Understanding a spoken or written paragraph is then a matter of habits, connections, mental bonds, but these have to be selected from so many others, and given relative weights so delicately, and used together in so elaborate an organization that "to read" means "to think," as truly as does "to evaluate" or "to invent" or "to demonstrate" or "to verify."

Authors:

Noam Chomsky
E. B. Huey
John Locke
B. F. Skinner
E. L. Thorndike

Figure 7. Matching Test 1.

Please read the five definitions of comprehension (understanding), and decide which of the five authors listed below them is being quoted in each.

1. Understanding is a constructive process, in which a representation is developed for the object that is understood. The difference between understanding and not understanding is in the nature of the representation. When a sentence is understood, its internal representation shows what the sentence means. The meaning corresponds to a pattern of relations among concepts that are mentioned in a sentence, and understanding is the act of constructing such a pattern.

2. The comprehension process is the mental operations which take place in the reader's head while he is reading. These operations are generally not observable and not open to introspection. On the other hand, the products of the comprehension process are the behaviors produced after comprehension has taken place, such as the answers to test questions.

3. In my opinion, comprehension is an almost perfect example of a gestalt, a total that is greater than the sum of its parts. It is undoubtedly true that the factors of word meanings, interrelationships of details, and reasoning are significant components of comprehension. These factors are recognized in a majority of the factor analyses of reading tests. Yet, certainly comprehension is more than these three simple elements, for this information leaves unanswered the questions of (a) what thinking processes operate in comprehension and (b) how these processes may be measured or trained.

4. In other words, during reading coded audio-visual and kinesthetic impressions derived from the descriptions of concrete objects are reassembled in the mind--this is comprehension.

5. The act of comprehending a sample of verbal material (a "message") consists, at least initially, of driving a "meaning" or "semantic interpretation" for it. Once the receiver of the message has derived this semantic interpretation, he may evaluate it for its "acceptability" to him (in terms, for example, of truth, relevance, or conformity to expectation), and if it is "acceptable" he may assimilate it to his cognitive structure, in which case we may say that he has "learned" the content of the message. In addition, he may derive further cognitive structure from the text on the basis of inferential processes . . . .

Authors:

John B. Carroll
James G. Greeno
Jack A. Holmes
Herbert D. Simons
George D. Spache

Figure 8. Matching Test 2.

Not because the definitions are necessarily better, or clearer, or even, I suspect, because you have ever actually seen any of them before. No, rather, because you know something about the author and his view of the world--in current parlance, his "schema." (And, perhaps, by elimination also, as test-wise readers.) My points, again, are simple ones.

1. There are many different conceptions of what comprehension is and very little commonality among them.

2. When, as today, there is so very little agreement, definers fall back on their own conceptions of the world--and make their definitions fit--rather than try to find common ground.

What kind of state of affairs is this for a science? It reminds me of Joel Greenspan's apt comment (I believe it was his): "In the physical sciences one stands on the shoulders of those who went before; in the social sciences, one steps in the face of those who went before."

We clearly need better agreement on such issues as:

1. A rough definition, at least, of reading comprehension.

2. How best, even though imperfectly, to measure reading comprehension.

3. What "level" of comprehension is desirable for the many different kinds of reading behaviors (e.g., leisure reading, school reading, and following directions).

The state of confusion in the area of comprehension clearly affects the area of readability. For example, different readability formulas predict anywhere from 50 percent to 100 percent comprehension of the McCall-Crabbs Test Passages, depending upon which one you choose. They also, in other cases, predict anywhere from 35 percent to 55 percent comprehension on an entirely different measure--cloze procedure. Yet, as far as most users are concerned, readability formulas are thought of as doing the same thing--predicting "comprehension."

The problem, then, is more that of an inability to agree on measures and levels of comprehension than it is on what should be included in predictors of readable writing. Similarly, until we know better what goes into comprehension, we can hardly be expected to help would-be producers of readable writing very much.

Now, since I have presumed to suggest what is needed in the areas of readability and comprehension, perhaps I had better get back to work on and with a possible framework for the study of readability.

Answer to Matching Tests 1 and 2:

Test 1	Test 2
1. Skinner	1. Greeno
2. Locke	2. Simons
3. Huey	3. Spache
4. Chomsky	4. Holmes
5. Thorndike	5. Carroll

THE PRACTICAL APPLICATION OF READABILITY/COMPREHENSIBILITY RESEARCH TO THE PRODUCTION OF MORE USABLE MILITARY TECHNICAL MATERIALS

J. D. Kniffin
Westinghouse Electric Corporation
Technical Logistics Data
Hunt Valley, MD 21030

ABSTRACT

This paper discusses many of the major products of recent readability/comprehensibility research conducted at the Westinghouse Electric Corporation. The paper describes practical tools and methods for applying the Westinghouse research and that of others to the production of more usable military technical materials. Major topics include methods for producing readable writing, use of readability formulas to predict the difficulty of text, methods of performing automatic readability calculations, the new military specification for readability, and the development of a computerized readability editor. The paper concludes with a discussion of some areas in which further research is required.

Introduction

Recent studies by all three military services have shown that technical materials are written at a level of difficulty well above the reading ability of the personnel who must read those materials (Kincaid, 1967; Smith & Kincaid, 1970; Klare, 1963; Caylor, Sticht, Fox, & Ford, 1973; Duffy & Nugent, 1974). Several researchers have indicated that the average reading ability of enlisted personnel is at about the ninth grade level. A recent GAO study (FPCD-77-13, 1977) strongly implies that within selected groups of servicemen, the average reading level may be well below the ninth grade level. It has also been found that technical manuals and training materials are often written at college level or beyond. Clearly then, a mismatch exists between the reading ability of the average serviceman and the readability of technical materials.

The causes of the reading ability-readability mismatch are basic. Modern military equipment is vastly more complex than that of just a few years ago, and this factor alone causes a problem. The problem is magnified when a writer is a specialist or expert and his readers are relative novices. Another difficulty is that the average reading ability of high school graduates is lower than it was 10 years ago. These considerations, in conjunction with the all volunteer concept in the military services, make selection of personnel with high reading ability progressively more difficult.

When the gap between reading ability and the difficulty level of writing exceeds approximately two grade levels, serious inefficiencies result. Reading speed, comprehension, and retention are all reduced. Among the practical consequences of this reading ability-readability mismatch are:

1. Interference with training of military personnel.
2. Errors in following technical directives.
3. Costly errors in performing equipment maintenance.
4. Increased down-time of complex expensive equipment.
5. Failure of technicians to use available technical manuals.

Surely, these consequences are grave. The military services cannot operate efficiently unless means are found to reduce the mismatch between the reading ability of servicemen and the readability of technical manuals.

There are two basic methods of solving the mismatch problem. The first is to improve the reading ability of technical manual users either by remedial instruction or by selection of personnel with high reading ability. Remedial instruction requires a great deal of time and effort before significant gains in reading ability can be achieved. As mentioned, selection of personnel with high reading ability is currently an extremely difficult task. Because of these constraints, a more practical solution to the mismatch problem is needed.

The other basic method of solving the problem is to improve the readability of technical materials. Though not an easy task, practical experience has shown that almost any technical manual can be made more readable. The problem, of course, is to produce readable materials without a loss of technical accuracy--a problem that can be solved.

To make sure that the manuals they purchase are readable, the military services have begun to impose readability requirements in the specifications for new manuals. The challenge facing Westinghouse Technical Logistics Data (TLD) and other suppliers is to make certain the technical manuals they produce meet the new readability specifications.

To meet this challenge, TLD began its own independent research program to study readability. Though a great deal of readability research had already been done, how much of that research was applicable to technical writing? Are there any significant differences between technical writing and traditional adult literature? How can a technical writer be sure his writing is at the appropriate level of difficulty? What is the most efficient way to estimate the reading grade level of text? If text is too difficult for intended readers, what research-proven techniques should be used to improve readability? Finding answers to these questions have been the major goals of TLD's readability research.

Throughout its research program, TLD has sought to develop practical tools and methods for applying both its own research and that of others. The balance of this paper will address the products of recent research at TLD and will describe ongoing programs. The paper will conclude with some observations about the current trends in readability/comprehensibility research and the need for future research.

Discussion

Development of Readability Guidelines

Approach. One of the major goals of the research was to develop readability guidelines for TLD writers. The purpose of the guidelines was to provide writers with research-proven techniques for producing readable writing and predicting the difficulty level of the text.

A preliminary investigation revealed that books designed to instruct writers in the techniques of clear writing would be of limited value in achieving the research goals. Some of the books were very interesting, and later research proved that a few of them were also quite accurate. The suggestions in most of the books were subjective, however, rather than objective in nature. Also, there was little consistency in the suggestions of various authors. Therefore, a thorough study of prior readability research into both traditional and technical writing was undertaken.

To begin with, an extensive study plan was developed. The plan outlined the course of the research effort and detailed over 50 factors that could conceivably have an effect on the communication of technical information. These factors included those related to technical concepts, readers, writers, and techniques of presenting technical information. No effort was made to prejudge the validity of these factors.

With the cooperation of Dr. George Klare of Ohio State University, a list of about 130 pertinent reference materials was prepared. The materials covered nearly all aspects of the "production" and "prediction" topics of readability research. Those materials that were obtained were organized into a formal readability library to support the research and to allow TLD writers to study readability on their own.

Using the study plan as an outline, the library materials were studied and critiqued over a period of several months. Information applicable to the readability of technical materials was then assembled and presented as readability guidelines.

Producing Readable Writing. The guidelines contain a number of suggestions for producing readable writing. These suggestions are based solely upon research findings and are therefore limited to word and sentence variables. So far, research of other variables such as paragraph construction, organization, and emphasis has provided little useful data for writers.

The most significant findings concerning changes in word variables were found to be:

1. Use familiar or frequently occurring words.

2. Use short words instead of long words.

3. Use words with high association value (words that quickly bring other words to mind).

4. Use concrete words (those that arouse an image in the mind) instead of abstract words.

5. Use active verbs instead of nominalizations (nominalizations are usually verbs made into noun form).

6. Limit or clarify the use of pronouns and other anaphora (words or phrases that refer back to a previous word or unit of text).

These suggestions apply to both traditional and technical writing. The caveat concerning technical writing is that writers should avoid changing technical or special meaning terms. Better substitutes cannot usually be found for those terms. Instead, writers should provide definitions along with the terms if the terms are unlikely to be familiar to intended readers.

The most significant suggestions concerning changes in sentence variables are as follows:

1. Write short sentences and clauses.
2. Form statements instead of questions where possible.

3. Make positive instead of negative statements where possible.
4. Make statements in active instead of passive voice where possible.

The research supporting changes in word variables is much stronger than that for sentences; indeed sentence variables will require a good deal more research. There is little doubt, however, that readers usually find simple, declarative, positive, active sentences to be most readable.

In this paper, the suggestions for making changes in word and sentence variables were merely listed. For a detailed look at the research basis behind these findings see "A Manual for Readable Writing" (Klare, 1975). In the manual, Dr. Klare describes a systematic approach for applying these suggestions to produce more readable writing. The TLD readability guidelines contain the same suggestions as Dr. Klare's manual, but modifications were made to extend his approach to technical writing.

Practical efforts at TLD have shown that it is a difficult task to learn to apply all of the word and sentence variable changes listed above. Generally, a writer must consider word and sentence variables simultaneously, but he must be careful to avoid mechanical application of the suggestions. With practice and the use of special source materials, most writers soon find that they can produce more readable writing. The most valuable source materials that TLD has found for making word variable changes are: the Thorndike-Barnhart Comprehensive Desk Dictionary (Clarence Barnhart, Ed., 1958), Thorndike-Barnhart Handy Dictionary (1955), Soule's Dictionary of English Synonyms (Alfred Sheffield, Ed., 1959), Computational Analysis of Present-day American English (Kucera & Francis, 1967), and the Living Word Vocabulary (Dale & O'Rourke, 1976).

Predicting the Reading Grade Level of Technical Material. Predicting the difficulty (reading grade level) of text, though not easy to do well, is relatively simple compared to producing readable writing. Of several methods for judging difficulty, the most convenient and specific method is to use a readability formula. Formulas are developed by studying the relationship of style variables in passages of text versus the test scores of readers taking comprehension tests on the same passages. There are an enormous number of style variables. Studies show, however, that proper counts of two simple variables provide as much or more predictive power than any of the complex variables.

These variables are average word length in syllables and average sentence length in words. When analyzing text, these variables are calculated and inserted into the readability formula. The formula is then calculated to yield the measure of difficulty, usually expressed as a reading grade level.

Hundreds of readability formulas have been developed. Some were designed for children's material; some, for adult material; some, for technical material, etc. There are only five or six formulas appropriate for use with military technical manuals. Of these, all but one has been developed within the last 6 years. For many complex reasons, the formula recommended for use by TLD writers is:

Grade Level = .39 (AVG. No. Words/Sentence) + 11.8 (AVG. No. Syllables/Word) - 15.59.

This formula, the Recalculated Flesch Reading Ease Formula, was developed for the Navy (Kincaid, Rogers, Fishburne, & Chissom, 1975) along with two other formulas. Although the Recalculated Flesch formula provides about the same readabiity scores as the two other formulas developed by Kincaid et al., several factors make it the appropriate choice of the three. Kincaid's Recalculated Flesch formula was also found to be more appropriate than the FORCAST formula (Caylor, Sticht, Fox, & Ford, 1973), and the RIDE SCALE (Carver, 1974). The reasons for selecting the Recalculated Flesch formula will be discussed in the next section.

To apply the formula, a writer must count the number of syllables, words, and sentences in the passage being analyzed. For long passages, several 200-word samples are chosen to save time. The formula variables and score are computed. A prediction of how readable the piece of writing is likely to be for intended readers is provided by the formula score. For example, if the grade level score is 12.7 but the intended readers average only 9th grade ability, the passage is likely to be too difficult. The writer should then rewrite the passage to suit his intended readers. A rewrite would, of course, also be needed if readability requirement specifications have not been met. After rewriting, the formula is applied again to verify that the passage is at the appropriate level.

Hand calculation of readability formulas is time consuming and, on a large volume of materials, can be expensive. TLD anticipated that the military services would begin to impose readability requirements for new technical manuals. For this reason, research was directed at finding a more cost-effective method of performing readability calculations.

Automated Readability Calculations. All the manuscript of a technical manual must, of course, by typewritten. The first draft and later versions of most TLD manuals are typed into the DOCUMATE text processing system, which has computational capability. If DOCUMATE could perform readability calculations in conjunction with text processing, a number of important benefits would evolve. Writers would learn instantly if their draft manuscript met readability requirements. The formula calculations would be performed rapidly and at nominal cost. Automatic readability was clearly worthwhile, but much had to be learned to made it work.

A small number of other researchers have prepared computer programs for automatic readability calculations. Programming techniques under development at the Service Research Division of General Motors offered the most promise to meet TLD's highly specialized goals. The GM techniques, referred to as STAR (undated), were incorporated into a BASIC language program for evaluation.

The original data used in developing Dr. Kincaid's Flesch formula was obtained for the evaluation. Hand counts were made of the number of words, syllables, and sentences in the experimental passages supplied by Dr. Kincaid. The counts were compared to the counts made using the BASIC program. These comparisons, along with exhaustive studies of vowel combinations, word endings, punctuation problems, etc., continued for almost 1 year. The algorithms in the BASIC program were updated numerous times until desirable accuracy was achieved.

The successful development of the BASIC program demonstrated that accurate readability calculations could be performed by computer. Mr. J. H. Griffith and Mr. E. J. Pierce of TLD's DOCUMATE Center have converted the original BASIC program into the more efficient ALGOL language used by DOCUMATE. The ALGOL program has been used to analyze numerous sample passages totaling over 10,000 words. On these samples, the word count is 100 percent accurate; the syllable count, 99 percent accurate; and the sentence count, 97 percent accurate. These figures compare favorably with those of the other known computer programs and with hand counts made by analysts. One other feature of the program is a routine that calculates the average reading grade level of any number of passages. This feature is useful because the new military specification (to be discussed later) requires that manuals be written at some particular average reading grade level.

A sample printout from the ALGOL program is shown as Figure 9. The printout shows the text analyzed and lists the words of three or more syllables in the order that those words occur in the text. A writer can use the long word list to identify quickly those parts of the text that are most likely to be difficult for intended readers. The SUMMARY AND CALCULATIONS portion of the printout lists the raw data, formula variables, and the reading grade level according to the Recalculated Flesch formula. Using this data, a writer can tell at a glance if his writing is an appropriate level or whether military specification requirements have been met.

New Military Specification for Readability

In 1977, TLD completed a study contract for the U.S. Army called "Criteria for Improved Readability and Understanding." The earlier TLD research was directly applicable to each phase of the Army contract. Although a small contract in terms of funding, this contract will have a major impact on the procurement of all future military technical manuals. Quality assurance provisions developed under this contract have recently been published as the new readability requirements of MIL-M-38784A.

The first task under this contract was to select the readability formula that was most appropriate for determining the reading grade level of Army technical manuals. The five recent formulas considered potentially the most appropriate for military technical materials were: the RIDE scale, Carver (1974); the FORCAST formula, Caylor et al. (1973); the Recalculated Automated Readability Index (ARI), Kincaid et al. (1975); the Recalculated Fog Count, Kincaid et al. (1975); and the Recalculated Flesch Reading Ease Formula, Kincaid et al. (1975). Of these five formulas, the Recalculated Flesch Reading Ease Formula was selected for the reasons detailed below.

The RIDE scale was rejected because the materials on which it was developed were not sufficiently technical in nature; the scores it provides are merely broad (not specific) levels of difficulty; further, calculation of the formula requires counting all of the letters in every word of a sample of text, a very time-consuming and expensive process.

The FORCAST formula was developed using military technical materials with military personnel serving as subjects, but certain factors rendered this formula suspect. First, the formula has only a word difficulty factor; it contains no sentence difficulty factor. While this feature does make mechanical manipulation of readability scores more difficult (for example, by arbitrarily dividing long sentences in two), numerous other researchers have found that the addition of a sentence difficulty factor adds substantially to the ability of a formula to predict comprehension. Second, the range of accurate grade level scores predicted by FORCAST appears to be very small. The most difficult materials used in developing the formula were only about the 12th-13th grade level; thus, the accuracy of FORCAST scores above this level is highly questionable. On the other hand, FORCAST scores near the low end of its scale (grade level five) appear to be even more questionable. (To rate a FORCAST score of grade level 5, a sample passage must contain 150 consecutive one syllable words; locating even one such passage is virtually impossible.) Third, it is somewhat difficult to automate (computerize) FORCAST calculations.

The three formulas developed by Kincaid et al. were all based on the same military technical materials and military personnel serving as subjects. The materials used spanned a wide range of difficulty, from about the 5th to the 16th grade levels. The subjects used (Navy enlistees) were closely matched to the overall level of the entire population of Navy enlisted personnel. On a large enough sample, all three formulas provide about the same grade level scores. The Recalculated Flesch formula was felt to be the most appropriate of the three because, unlike the other two, it has all of the following attributes:

LONG WORDS:

velocity	important	consideration
formation	horizontal	produces
several	temperature	vertical
adiabatic	additional	saturation
additional	temporary	supersturated
producing	Adiabatic	usually
particles	particles	horizontal
visibility	However	particles
evaporation	horizontal	visibility
restricted	moderate	condition

SUMMARY AND CALCULATIONS:

NUMBER OF SENTENCES=	10
NUMBER OF WORDS=	198
NUMBER OF SYLLABLES=	330
AVERAGE SENTENCE LENGTH=	19.8
AVG SYLLABLES PER WORD=	1.67
GRADE LEVEL EQUIV=	11.8

Wind velocity is an important consideration in the formation of fog and/or low ceiling clouds. The horizontal motion of the air next to the earth's surface produces friction which, in turn, causes the air near the ground to tumble, setting up eddy currents. The size of the eddy currents vary with the wind speed and the roughness of the terrain. Lower wind currents produce more shallow eddies, and stronger wind currents produce eddies up to several hundred feet and higher.

When the temperature and dewpoint are close at the surface and eddy currents are 100 feet or more in vertical thickness, adiabatic cooling in the upper side of the eddy could give the additional cooling needed to bring about saturation. Any additional cooling would place the air in a temporary supersturated state. The extra moisture will then condense out of the air, producing a low ceiling cloud. Adiabatic heating on the downward side of the eddy will usually dissolve the cloud particles. If all cloud particles dissolve before reaching the ground, the horizontal visibility should be good. However, if many particles reach the ground before evaporation, the horizontal visibility will be restricted by a moderate ground fog condition.

Figure 9. Computer print-out of readability analysis.

1. Flesch-type formulas have been used extensively for many years and thus a greater number of people are likely to be familiar with their application. (A survey by the author of 328 studies on the application of readability formulas revealed that Flesch-type formulas are commonly used with educational materials. No instance could be found, however, where any of the other formulas considered had been similarly applied.)

2. The counts required for Flesch-type formulas (sentences, words, and syllables) are easy to make: Extraordinary analyst skills are not required. Studies have shown that Flesch-type formula scores calculated on the same passages were consistent when calculated by different analysts.

3. The counts required for Flesch-type formulas can be made manually or automatically without too much inconvenience. Fog counts are quite awkward on single-spaced type, and the counts for the ARI are usually done only by computer or a specially adapted typewriter. The ability to make calculations either by hand or computer also meant that no contractor would be forced to invest in a software development program.

After the Recalculated Flesch formula was selected, the next task under the Army contract was to apply the formula to about 100 randomly selected samples in four Army manuals. Many types of technical writing, including introductory materials, descriptive materials, operational instructions, etc., were analyzed. Three of the most difficult samples (highest formula scores) from each manual were then rewritten for readability. Six samples were rewritten to about the 5th grade level; three, to the 6th grade level; and three, to the 7th grade level. Army personnel verified that the meaning of the rewritten samples was not changed during the rewriting process.

Obviously, not all technical materials can be rewritten to the low readability grade-level scores such as was done on this contract. (None of the manuals supplied by the Army covered sophisticated electronic systems, for example.) This contract did demonstrate, however, that a great deal of text simplification can be made to almost any type of technical writing.

The last major phase of the Army contract was to develop quality assurance provisions to ensure that the readability of technical manuals is appropriate for intended users. It was these quality assurance provisions that have now become the standard for all military technical manuals. The attachment to this paper is a copy of the pertinent portions of Amendment 5 to MIL-M-38784A. The paragraphs below briefly describe some major features of the specification.

The specification provides that the procuring activity must establish the appropriate reading grade level for each manual. The major reason for this provision is simply that the military services themselves are the only source of either accurate or up-to-date data concerning the reading skills of intended users. In choosing the appropriate reading grade level for a manual, the services must be both objective and fair to contractors. It was in no way intended, for example, that the Navy should specify that all manuals be written to the average reading ability of Navy enlistees. It is simply not necessary (or worth the expense) to write 9th grade level manuals for intended readers, such as pilots or radar operators, with high level reading skills. What was intended was that all the services establish target grade levels based on the actual needs of specific groups of personnel within each service. The Air Force has already established a target grade level for written materials used in each of 250 job specialty codes. It is hoped that the other services will follow the lead of the Air Force and establish specific target grade levels for their personnel.

The specification also provides that the overall grade level of each manual, as calculated by the Recalculated Flesch formula, must be no more than 1.0 grade levels above the appropriate level. Each sample within a manual must be no more than 3.0 grade levels above the appropriate level. These tolerances were provided to assure that manuals are consistently near the reading ability of intended users. At the same time, the tolerances acknowledge that it may not be possible to write every part of a manual to a specific reading grade level, especially if that level is rather low.

A sampling procedure is also provided in the specification. Rules are provided for determining both the number of samples to be analyzed in a given manual and the size of each of those samples. Samples are selected at intervals throughout the text to assure adequate coverage of the entire manual. Every "Nth" page of text, as determined by a "look-up" table, is sampled. This feature was designed to eliminate any tendency to "randomly select" easy samples for analysis. The size of each sample, about 200 words, conforms to the size preferred by Dr. Kincaid in developing the formula.

Finally, the specification outlines the rules for counting sentences, words, and syllables. These rules basically conform to those presented by Kincaid et al. (1975), but there are two minor modifications. First, all numbers are counted as one syllable. Second, all acronyms and abbreviations are counted as one syllable unless they actually spell out a word of more than one syllable (ARMCOM, for example, is counted as two syllables). These rule changes are a compromise to the rules suggested by Flesch (1948). Rather than ignore numbers (when a passage contains a large amount of numbers) or place undue emphasis on the difficulty of acronyms, these rule changes account for numbers and acronyms in a manner that is more realistic for technical materials. These changes also facilitate automatic calculation of syllables.

The new specification became effective for all the military services on 24 July 1978. Again, it is hoped that all the other services will follow the lead of the Air Force in determining the specific needs of its personnel. It is also hoped that, by following the specification, contractors will produce manuals that are at an appropriate level of difficulty for intended readers.

Current TLD Research

Computerized Readability Editing. The research into automated readability calculations led to another intriguing thought. If the computer can locate long words and sentences, can it also be programmed to identify other style variables known to affect readability and comprehension? A computer program that edits for readability could be a very useful aid to TLD writers. If a writer was having difficulty meeting readability requirements, the computerized editor could suggest style changes to make his writing more readable. The writer could then concentrate his efforts upon very specific parts of his manuscript that may need revision.

Recent TLD research has identified several style variables that are candidates for computer identification. Current efforts are directed at choosing the style variables that can most readily be identified and selecting a format for presenting this data to writers.

One feature of the current research involves developing a computerized word and phrase "substitution dictionary." A computer program will automatically analyze text and flag long or unfamiliar words and phrases. For each word or phrase flagged, the program will provide a writer with a list of shorter, more readable, substitute words.

English as a Second Language. Another topic of current research is the study of English as a second language. At present, TLD produces technical manuals in English for several foreign customers. There is considerable business with such countries as Iran, Japan, West Germany, Venezuela, Greece, Brazil, and South Korea. While manuals for these countries are prepared in English, study is needed to determine the average English language ability of foreign technicians as well as idiosyncracies in their use of English. For example, if the average technician of a particular country reads English at only the 6th grade level, a manual written at the 11th grade level would be difficult or impossible for him to read. Also, English usage in various countries must be studied so that inadvertent maintenance mistakes may be prevented. For example, if a foreign technician would say "close the switch" to turn off a circuit but an American would say "open the switch," the foreigner would perform the wrong action because of the way he normally uses the English language. There are many other possible problem areas where English is used as a second language.

TLD is currently assembling information on the use of English as a second language in selected foreign countries. This information will be used to assess the general background of English language usage in the selected countries. Contractual support is being sought so the particular needs of foreign military technicians can be assessed. Hopefully, the results of this study will enable TLD to better serve its foreign customers by producing manuals that account for English language usage by those customers.

Some Observations About Needed Research in the Field of Readability/Comprehensibility of Technical Materials

A great deal of research has been done on the readability/comprehensibility of military technical materials. As in other fields, however, much work remains to be done. In addition, other work (which is not true research) is needed so the results of research can be better applied to the real-world situation. This paper will conclude with some observations about needed research in the field of readability/comprehensibility as it applies to military materials and personnel.

One way in which the military services could aid technical writers would be to provide more data on actual user populations. As mentioned, the Air Force has provided target reading grade levels for many of its occupational specialty codes, and, hopefully, the other services will follow this lead. There are other data, though, that could help writers form a mental picture of their intended readers. An educational profile of typical users may be useful. A description of the actual training of users may be useful. A precise job description of various specialty codes may be useful. An experience profile of actual users may be useful. This type of data could help writers prepare materials that are more appropriate for intended readers. For example, this data could help ensure that all pertinent materials are presented but that material beyond the training or job function of personnel will be excluded. Such "tailor-made" writing may well help readers maintain their motivation to use the materials provided for them.

Many areas of traditional readability research need to be expanded to cover technical writing. Some areas for needed research are listed below:

1. Perform format optimization studies so that various types of technical data are presented in the most useful forms.

2. Study the effect of the omission of articles (a common practice in technical writing) on the comprehensibility of text.

3. Study the effect of the use of acronyms and abbreviations (another common practice in technical writing) on the comprehensibility of text.

4. Develop concreteness ratings for technical terms (highly concrete words are those that easily arouse an image in the mind).

5. Study indexing and referencing schemes so the interaction between text and illustrations can be optimized.

6. Study emphasis techniques such as the use of underscoring, symbols, color, etc. to determine if these techniques improve comprehension.

Also, there is another major research topic upon which very little work has been done. That topic is the study of the comprehensibility of illustrations. Illustrations sometimes comprise 50 percent of a technical manual, yet little is known about the relative difficulty in understanding various types of graphics. In theory, illustrations, like text, can be either easy or hard to understand or to use. Subjectively, such factors as illustration type, size, placement, and information density should all contribute to comprehensibility. Can variables of graphics presentations be identified? If so, can these variables be measured, counted, or otherwise accounted for objectively? Can the relative difficulty of illustrations be measured objectively? Can formulas be developed for predicting the comprehensibility of illustrations? These questions cannot be answered at the present time, and, yet, illustrations are a very expensive way to convey information. Therefore, research that will lead to a better understanding of how to produce comprehensible illustrations is very much in order.

Finally, most studies of the effect of readability upon comprehension have been controlled laboratory studies. In such studies, subjects typically read specially prepared materials (altered for readability) and then take reading comprehension tests on those materials. Klare (1976) has expressed some concern about the question of generalizing the results of such studies to a real-world language population. Much of this concern is due to the fact that the motivation of subjects to do well in a test situation often obscures the effect of readability on comprehension. Both Dr. Klare and this author feel that a more nearly ideal answer to the effect of readability must come from field studies using unobstrusive measures. In such studies, subjects are unaware that they are being tested. The test conditions then closely match typical levels of motivation and typical conditions of study.

Either training materials or job materials could be used in field studies using unobtrusive measures. In either case, one group of subjects would use materials revised for readability while a control group of subjects would use existing materials with no revisions. The two groups would be matched in terms of reading ability. If training materials were used, some possible unobtrusive measures would be the (1) length of time taken by subjects to complete course materials, (2) percentage of subjects who complete a course of study, (3) percentage of subjects who pass a standard examination based on the training materials, and (4) average scores on standard examinations based on training materials. If job materials were used, some unobtrusive measures would be the (1) average length of time required to identify specific malfunctions in equipment, (2) average length of time required to repair specific malfunctions in equipment, (3) length of time required to complete routine job functions, and (4) percentage of mistakes made in using the job materials. While such studies will surely not be easy to perform, they will provide real-world facts concerning the effect of readability upon comprehension.

COMPUTER-STORED NAVY WORD LISTS FOR READABILITY ANALYSIS AND LITERACY TRAINING

J. Peter Kincaid

U.S. Navy Training Analysis and Evaluation Group
Orlando, FL 32913

ABSTRACT

This paper describes the ongoing development and future use of a series of word frequency lists. A general word list is being constructed by completing a frequency analysis of recruit training materials; additional supplemental word lists relevant to four clusters of Navy job specialties will also be constructed. Computer programs and procedures will be developed so that the lists can be used to (1) give feedback to the technical writer during the drafting of training materials and technical manuals, (2) form the major component of a readability formula, and (3) identify unfamiliar words that should be stressed in functional literacy training.

Introduction

The time is ripe to create a new word list of Navy training words that would have three major uses: (1) as the major component in a readability formula patterned after the Dale-Chall readability formula, (2) as a tool for creating functional vocabulary exercises for remedial reading instruction, and (3) to provide feedback to technical writers.

A recently completed review of current and proposed technical manual systems completed by Hughes Aircraft Company (1977) for the Naval Technical Information Presentation Program states:

> Writers (of technical manuals) need . . . vocabulary tools . . . and checks of readabililty of in process and completed draft technical manuals.

Curran (1977), in his careful review of the state-of-the-art in Navy readability, also stresses the need for vocabulary analyses of military training materials.

The most current word list derived from military training curriculum was published in 1964 by the American Institutes for Research. While the list was carefully compiled, it is now somewhat out of date. Other word lists are in wide use (e.g., the Dale-Chall Formula, 1948), but are employed for producing reading material for elementary school children rather than for military technical trainees.

Given that the Navy is moving toward the computer processing of manuals used for training (Keeler, 1977), it is desirable to provide feedback to the technical writer about the readability of his draft materials. This kind of feedback can be effectively given via a computer analysis. A vocabulary analysis of the draft materials would be very useful to the technical writer. Current Navy readability standards contain requirements for the use of formulas, like the Flesch Reading Ease formula (Flesch, 1948), which contains a measure of word length and a measure of sentence length combined in a formula to predict grade level of reading difficulty. Vocabulary controls are not a part of the current standards with the exception that MIL-M-81927(AS), 15 February 1975, provides for the use of a "list of preferred verbs," This is a list of 278 verbs recommended for inclusion in training documents used by the Naval Air Command.

The Training Analysis and Evaluation Group (TAEG) has been tasked by the Chief of Naval Education and Training to develop a dictionary of most commonly occurring words in the recruit training curriculum and additional lists for selected clusters of rates (i.e., Navy occupational specialties).

Issues in Creating the List

The development of a Navy word list format requires:

1. Identification of the user population (i.e., the particular technical specialists for which the lists are intended and their levels of experience).

2. Establishment of appropriate procedures for sampling training materials and technical materials in creating the word lists.

3. Design of automated strategies for using the word lists to provide feedback to technical writers during the various phases of writing the materials.

4. A decision about how long the lists should be.

Four existing word lists provide insight into the design of the Navy word lists.

1. The Dale-Chall list (1948) is probably the most widely used and validated of existing word lists. It is also the principal component in the Dale-Chall readability formula, which is used to grade elementary school texts by most major educational publishing houses.

2. The Harris-Jacobson list (1975) was developed with heavy reliance on computer processing of text. The list is appropriate for a wider range of grade levels than the Dale-Chall word list but, like the Dale-Chall list, its primary purpose is for use with elementary school materials. The list also forms the principal component of a readability formula.

3. The Kucera-Francis list (1967) was developed using a large and varied sample of adult reading materials. The list is long, containing thousands of words appearing only once in the sample. It provides valuable information about the linguistic structure of the English language, but was not designed as a readability tool.

4. A list of words most commonly encountered in military training materials was compiled by the American Institutes for Research under Contract to the U.S. Navy in 1964. Although somewhat out of date today, the word list (containing 1745 words in rank order) has a content that is closer to the projected content of the Navy lists under development than any other existing word list. It is not the basis of a readability formula, however, and is a general word list having no associated technical word lists. Since it was manually compiled, it was not specifically constructed for computer use. The procedures for its construction do not provide a good model for the current effort.

Characteristics of Navy Word Frequency Lists

Inflected Endings

The Harris-Jacobson word lists provide the best model for the development of Navy word lists. Only root words are included in the list. The rules for this are contained in Table 5 below. The Navy word lists will also treat most variations from the root word as being equivalent to the root word.

Table 5

Rules for Root Words[a]

Root word plus	-s (plural), -y, -ly, -ily -s, -es, -'s (possessive) -d, -ed, -er, -est (comparative)
All words with double consonant before	-ing, -er (comparative), -est
All words dropping final -e before	-ed, -ing, -er (comparative), -est
All words changing y to i before adding	-ed, -es, -er (comparative), -est

[a]Rules used by Harris and Jacobson (1975) in developing their word lists. Each variation is considered equivalent to the root word.

General Word List

The current plan is to develop five word lists including a general word list and four supplementary technical lists. The general word list will be relevant to Navy recruit and apprentice training and will be based on virtually the entire written recruit training curriculum as taught at the Recruit Training Center in Orlando, Florida. The two major written documents in the recruit training curriculum as taught at all three Navy recruit training centers (located in Orlando, FL, Great Lakes, IL, and San Diego, CA) are the Bluejackets' Manual, soon to be published in the 20th Edition, and the Rate Training Manual (NAVTRA 10054-D), Basic Military Requirements, (1973). Not all chapters in these books are used in recruit training. The general word list will be compiled from about 300,000 words that are included in the curriculum.

The text of the Bluejackets' Manual is available in machine readable form, which means that it is nearly ready for the computer word frequency analysis. The text of Basic Military Requirements will have to be keyboarded before processing on TAEG's Wang computer. Proper nouns, most abbreviations, and numbers will be excluded from the list as in the Harris-Jacobson word lists.

Length of General Word List

Figure 10 illustrates the point that fewer different words account for a given proportion of words in military training materials than in general adult reading materials. For example, the 1000 most commonly occurring words in the military training curriculum accounted for nearly 90 percent of the total word count. The 1000 most commonly occurring words in a sample of adult reading materials accounted for less than 75 percent of the total word count.

The list will consist of from 1500-2000 root words representing about 90 percent of the total word count found in the curriculum of nontechnical military training courses. The American Institutes for Research military word list consisted of 1745 words, which accounted for nearly 93 percent of the total word count of the sample used in compiling the list.

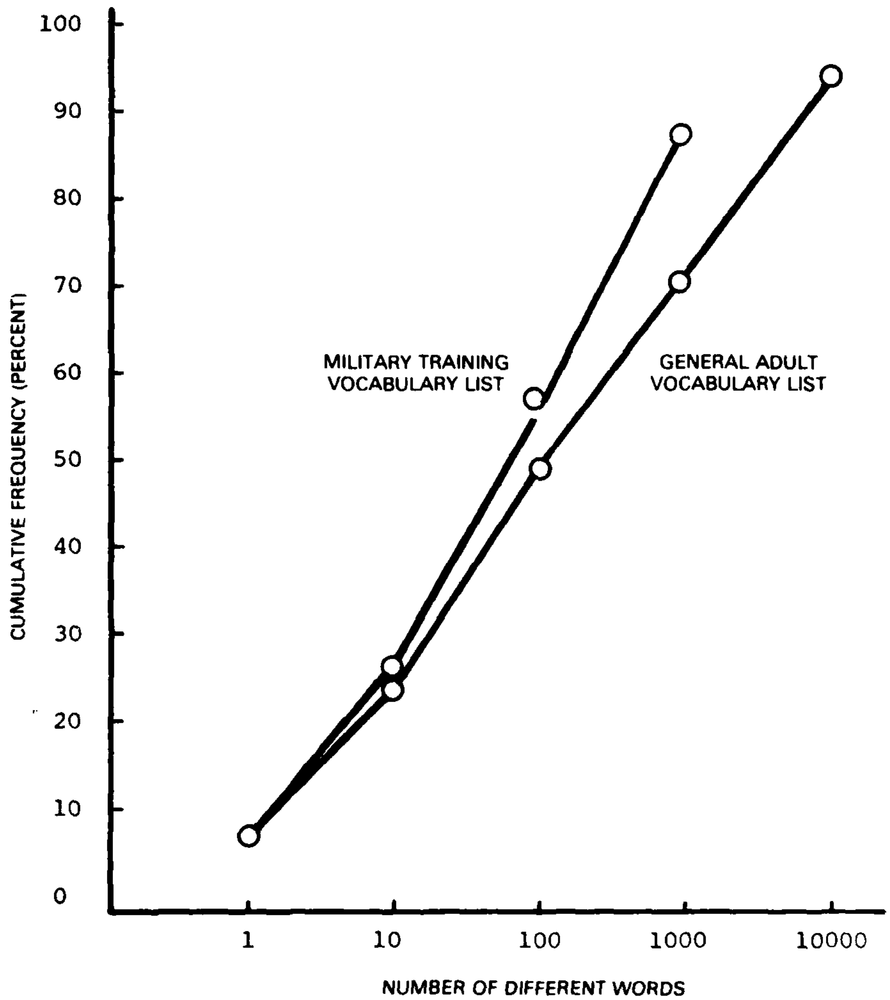


Figure 10. Comparison of the cumulative word frequency of a general adult vocabulary list (Kucera & Francis, 1967) and a military training vocabulary list (American Institutes for Research, 1964). Fewer different words are used in military training material than in general adult reading material.

Supplemental Word Lists

Supplemental technical word lists are needed that take into account the variety of different technical words found in different technical fields in the Navy. For example, certain technical words that commonly occur in reading materials used by an electronics technician are very different than technical words used by a personnel clerk. Separate technical word lists could be devised for a wide variety of technical specialties or for personnel having different levels of experience. Such a proliferation of word lists could easily become unwieldy.

Fortunately, the Navy has recently identified four functional clusters of job specialities (rates) for the Job Oriented Basic Skills (JOBS) program (see Table 6). The four cluster areas include propulsion engineering, a lower level electronics cluster that is labeled as "Electronics I," a more advanced electronics cluster that is labeled as "Electronics II," and a combined administrative and clerical area. The first level Navy schools listed (called "A" schools), which train specialists, account for over 50 percent of the total number of "A" school graduates in the Navy. These job clusters identify the four areas for which technical supplemental word lists need be developed. It may be desirable to develop only one technical word list relevant to both of the electronics clusters.

Table 6

Functional Clusters of Rates for JOBS[a] Program

Cluster	Rate/School
Propulsion Engineering	Boiler Technician (BT) Engineman (EN) Machinist's Mate (MM)
Electronics I	Gunner's Mate (GM) Missiles (GMM) Guns (GMG) Anti-Submarine Rocket (ASROC) Missiles (GMT)
Electronics II	Electronics Technician (ET) Radar (ETR) Navigational Equipment (ETN)
Administrative/Clerical	Yoeman (YN) Personnelman (PN) Storekeeper (SK)

[a]JOBS stands for "Job Oriented Basic Skills" and is a Navy program to develop functional reading skills for various Navy jobs.

Unlike the development of the general word list, in which the entire curriculum is being processed for the word frequency count, the development of each technical word list will require a sampling procedure. Curriculum material will be sampled from each of the technical schools listed in Table 6. Each list will require the processing of approximately 200,000 words. Most of the frequently occurring words will also occur on the general word list. Only technical words will be retained on the supplemental technical word lists. These supplemental word lists should be short, containing no more than about 200 words.

Development of Readability Formula Based on Word Lists

Once the word lists are developed, a readability formula should be developed incorporating the word lists as the major component. Such formulas are the best way to assess the readability of military training materials and technical manuals. The Navy readability formula would be patterned after the Dale-Chall or Harris-Jacobson readability formulas. These formulas have as their measure of word difficulty the percentage of words in a sample of text not found on the appropriate word list. The formulas also contain a measure of sentence difficulty--sentence length.

In the case of Navy technical training material to be graded for readability, the material would be compared against a composite word list consisting of the general word list and the appropriate supplemental technical list. In the case of general materials not containing a preponderance of technical words, only the general word list would be used.

There are several alternatives in the choices of an appropriate readability formula. As suggested by Stocker (1971, 1972) and Curran (1977), it is possible to use an existing formula, simply by substituting the new Navy word lists for the existing word lists. Either the Dale-Chall or Harris-Jacobson formulas are the most likely candidates if this procedure is followed.

Another, more valid, approach is to use a set of appropriate passages that have been tested for grade level of reading difficulty and derive a new formula. Kincaid, Fishburne, Rogers, and Chissom (1975) used 18 rate training manual passages to derive a modification of the Flesch reading ease formula. These passages were scaled for reading difficulty level by having Navy personnel, who had been tested for their reading ability, take a comprehension test on the passages. Thus, the grade level of the passages was determined by having Navy personnel, with known reading abilities, read and understand Navy materials. These passages could serve to derive the new word list formula but they are not ideally suited for this; the total number of words in the 18 passages was only about 3000 words and they were not specifically selected to reflect the major technical areas of the Navy rates. This was not necessary as the word difficulty measure in the modified Flesch readability formula was word length rather than vocabulary.

The most valid approach is to derive a new formula, carefully selecting the passages to reflect the technical areas of the supplemental technical word lists as well as general content. Navy enlisted personnel would serve as subjects with specialists in the several technical areas being tested with passages from their own particular specialty as well as passages having a general content. Given this stratified sample, it would be necessary to test more subjects than were tested in the previous effort, which utilized a little more than 500 subjects. Scaling the passages for level of reading difficulty is a major undertaking. As in the previous study, the formula would be derived using a computer to compute the multiple regression equation. This is a simple procedure once the passages have been scaled for grade level of reading difficulty. Each passage would be evaluated against the appropriate word list to determine the percent of words in the passage that are not included on the lists. Average sentence length of each passage would also be measured. Then these two predictors of reading difficulty, together with the criterion measure for each passage (scaled level of reading difficulty), would be analyzed in a computer to derive a multiple regression equation (or series of equations) of the form:

$$\text{Grade level} = B_1 \text{ (\% of words not on list)} + B_2 \text{ (average sentence length)} - C,$$

where B_1 and B_2 are weights for the two formula factors and C is a constant.

Uses of the Word Lists

TAEG has the capability to make selected tests of the word list and readability formulas. The word lists and associated programs for their use would also be made available to other interested agencies. TAEG has an ongoing project in computer-aided authoring (Braby, Parrish, Guitard, & Aagard, in press; Guitard, in press). This project is jointly sponsored by the Chief of Naval Education and Training and the Naval Technical Information Presentation Program. We have plans to build the routines for vocabulary and readability analysis into the computer-aided authoring system. TAEG could also provide a vocabulary and readability analysis of a few draft training or technical manuals provided their text is already available in machine readable form. The object would be to test and refine the techniques and to demonstrate their utility to potential users.

There are three potential areas of use for the word lists:

1. The list should serve as part of a computerized program to provide guidance to technical writers during the drafting process. The program would identify those words not on the list of common words and calculate average sentence length and average word length. It could include additional dictionaries such as a list of preferred verbs and the printout could suggest the preferred verb if a nonpreferred verb were to be included in the initial draft. Actually, a whole series of aids could be provided to the technical writer if the program were to include some of the features of the CARET I program (Klare, Rowe, St. John, & Stolurow, 1969). This program prints out the number of syllables under each word, notes unusually long sentences, and calculates several readability formulas. Curran (1977) gives additional suggestions about providing feedback to the technical writer concerning the readability of draft materials.

2. The lists should be used in a readability formula to assess the reading difficulty level of training materials and technical manuals. Various ways of doing this are discussed above. Once an appropriate formula is available, the question arises as to whether the formula should be made a standard for the development of Navy training materials and technical manuals. Currently, military standards call for the use of formulas using syllable length as the word difficulty factor, such as the Flesch Reading Ease Formula (Flesch, 1948), the revised Flesch formula (Kincaid et al., 1975), or the FORCAST formula (Caylor, Stitch, Fox, & Ford, 1972). A formula incorporating an appropriate word list is clearly a better readability measure, but it is also much harder to apply, virtually requiring a computer. It may not be reasonable to impose this requirement on contractors that write technical manuals when some do not have ready access to computer processing of the text of the manuals.

3. The lists will have several applications for literacy training. The lists themselves could form the basis of vocabulary exercises for enlisted personnel improving their low reading abilities in remedial reading classes. Since the lists will be based directly on Navy reading materials, learning the words will help the enlisted man perform essential job reading tasks. Programs containing the lists can also be used to identify unfamiliar words in existing training materials and technical manuals. This should aid in the construction of glossaries and other reading aids for the man on the job.

VALIDATION OF THE NAVY READABILITY INDICES

Robert P. Fishburne, Jr.

Calspan Corporation
Advanced Technology Division
Norfolk, VA

ABSTRACT

A study was designed to validate the Navy Readability Indices (NRIs) in an individualized instructional system using the criteria of comprehension and learning time. Two hundred Navy enlisted personnel enrolled in a computer-managed technical training course were tested for both comprehension and learning-time-to-criterion on four programmed instructional modules. The subjects were divided into four groups having reading ability grade levels (RGLs) of 9.3, 10.1, 10.8, and 11.7. The models were classified by modular readability grade levels (MGLs) of 9.1, 10.6, 11.3, and 12.5 using the NRIs. Results, which were supported by a replication, indicated that both readability and reading ability were significant predictors of comprehension and learning-time-to-criterion. In addition to validating the NRIs for programming instructional materials, the study has implications for further research into the learning time criterion for readability.

Introduction

The development of the Navy Readability Indices (NRIs) involved the recalculation of three readability formulas to make them more suitable for military use (Kincaid et al., 1975; Kincaid & Fishburne, 1977). The three formulas are the ARI, the "Fog Count," and the Flesch Reading Ease Formula. They were derived from test results of 531 Navy enlisted personnel enrolled in four technical training schools at two Navy bases. Personnel were tested for their reading comprehension level according to the comprehension section of the Gates-McGinite reading test. At the same, time, they were tested for the comprehension of 18 passages taken from rate training manuals. The average general classification test (GCT) score of the sample (54.7) was very close to the average GCT of the entire population of Navy enlisted personnel (about 54), indicating that the results should generalize to the entire population of Navy enlisted personnel as well as to military enlisted personnel in general. Scores on the reading test and training material passages allowed the recalculation of the grade level of the passages. This scaled reading grade level is based on military personnel reading military training material and comprehending it. Thus, the three recalculated formulas (derived using multiple regression techniques) are specifically for use with military training materials. Furthermore, the formulas are interchangeable because they were all calculated using the same data base.

The purpose of this paper is to present a study on the validation of the NRIs within an individualized instructional system using the criteria of comprehension and learning time. These criteria are operationally defined as percent error scores on modular comprehension tests, and minutes per learning objective to criterion values.

While precision in discriminating between comprehension levels is a worthy goal for readability validation research in itself, the addition of the learning time criterion has particular importance. The special significance of learning time in the self-instructional and computer-managed instruction (CMI) setting rests in the benefits derived from increased cost effectiveness with reductions in learning times. It must be recognized that applied uses of this measure must await rigid experimental evaluation with passages rewritten to specific levels of readability and analyzed through multiple regression or other appropriate statistical procedures. This investigation has been an attempt to establish only the basic relationship between learning time and readability.

Approach

The performance of 200 technical training students in the Navy's CMI system was monitored for both comprehension and learning time as they progressed through an ordered sequence of linear-programmed instructional materials. Percent error scores on objective multiple-choice tests and time-to-criterion data for these subjects were collected on four instructional modules representative of the range of readability within the curriculum. This data represented the modular readability grade levels MGLs of 9.1, 10.6, 11.3, and 12.5 as measured by the recalculated version of the Flesch Reading Ease Formula (i.e, .39 (words/sentence) + 11.8 (syllables/word) - 15.59). These data were organized into four groups of equal size according to each subject's position on a continuum of general classification test (GCT) scores ranging from a low of 41 to a high of 74. These GCT scores were then transformed to group mean reading grade level (RGL) designations of 9.3, 10.1, 10.8, and 11.7 by the Navy's GCT to reading level conversion formula: Grade Level = 2.7 + .14 GCT.

A two-factor (A x B) experimental design was utilized in which each of the modules classified by readability grade levels (A treatments) in combination with any one student reading grade level (B treatment) was administered to the same subjects, but with each B treatment characterizing a different group of subjects. The subjects were divided into four groups of 50 each. The total experiment may be regarded as having consisted of four treatment x subjects experiments. Thus, the A effect (readability) was a "within" subjects effect and the B effect (reading ability) was a "between" subjects effect.

Using this design and procedure, a replication of the study was conducted on a secondary sample of 100 subjects on four additional programmed instructional modules from the same curriculum.

Results

Table 7 presents the group means from the primary statistical analysis of comprehension test scores at each of four reading grade levels (RGLs) and four modular readability grade levels (MGLs). The percent error scores for Ss first attempt in each modular test ranged from a high of 10.26 for the poorest readers (RGL = 9.3) to a low of 4.53 for the best readers (RGL = 11.7). Likewise, the percent error scores ranged from a high of 13.06 for the module with the highest readability grade level designation (MGL = 12.5) to a low of 2.66 for the module with the lowest readability grade level designation (MGL = 9.1).

Table 7

Mean Comprehension Test Scores (Percent Error) for Levels of Reading Ability and Readability

	RGL = 11.7	RGL = 10.8	RGL = 10.1	RGL = 9.3
Reading Ability	4.53	8.78	7.96	10.26
	MGL = 9.1	MGL = 10.6	MGL = 11.3	MGL = 12.5
Readability	2.66	5.19	10.63	13.06

Figure 11 displays the comprehension test scores (percent error) for each of the four reading grade levels of subjects as a function of modular readability. These plotted cell means generally appear to increase with readability grade level and to decrease with reading ability. The differences "between" (reading ability) and "within" (readability) factors and their interaction are significant at the $p < .001$, $p < .001$, and $p = < .05$ levels respectively.

Table 8 presents the group means for the primary statistical analysis of time-to-criterion scores at each of four reading grade levels and four readability grade levels. The values ranged from a high of 4.69 minutes per learning objective for the poorest readers (RGL = 9.3) to a low of 3.14 for the best readers (RGL = 11.7). Likewise, the time-to-criterion values ranged from a highest readability grade level designation (MGL = 12.5) to a low of 1.94 for the module with the lowest readability grade level designation (MGL = 9.1).

Figure 12 displays the learning time (minutes per learning objective to criterion) values for each of the four reading grade levels of subjects as a function of modular readability. In the same manner as the comprehension test score data (percent error), the plotted cell means for learning time are observed to increase with readability grade level and to decrease with reading ability. Significant differences "between" factors (reading ability), "within" factors (readability), and their interaction are indicated, all at the $p < .001$ level.

Replication Study

The primary results for both comprehension test scores (percent error) and learning-times-to-criterion (minutes per learning objective) were generally replicated. Significant differences were found at the $p < .001$ level between reading grade levels, but for the percent error data only. Significant differences "within" (readability), however, were found at the $p < .001$ level for both error data and learning time data. No significant interactions were indicated.

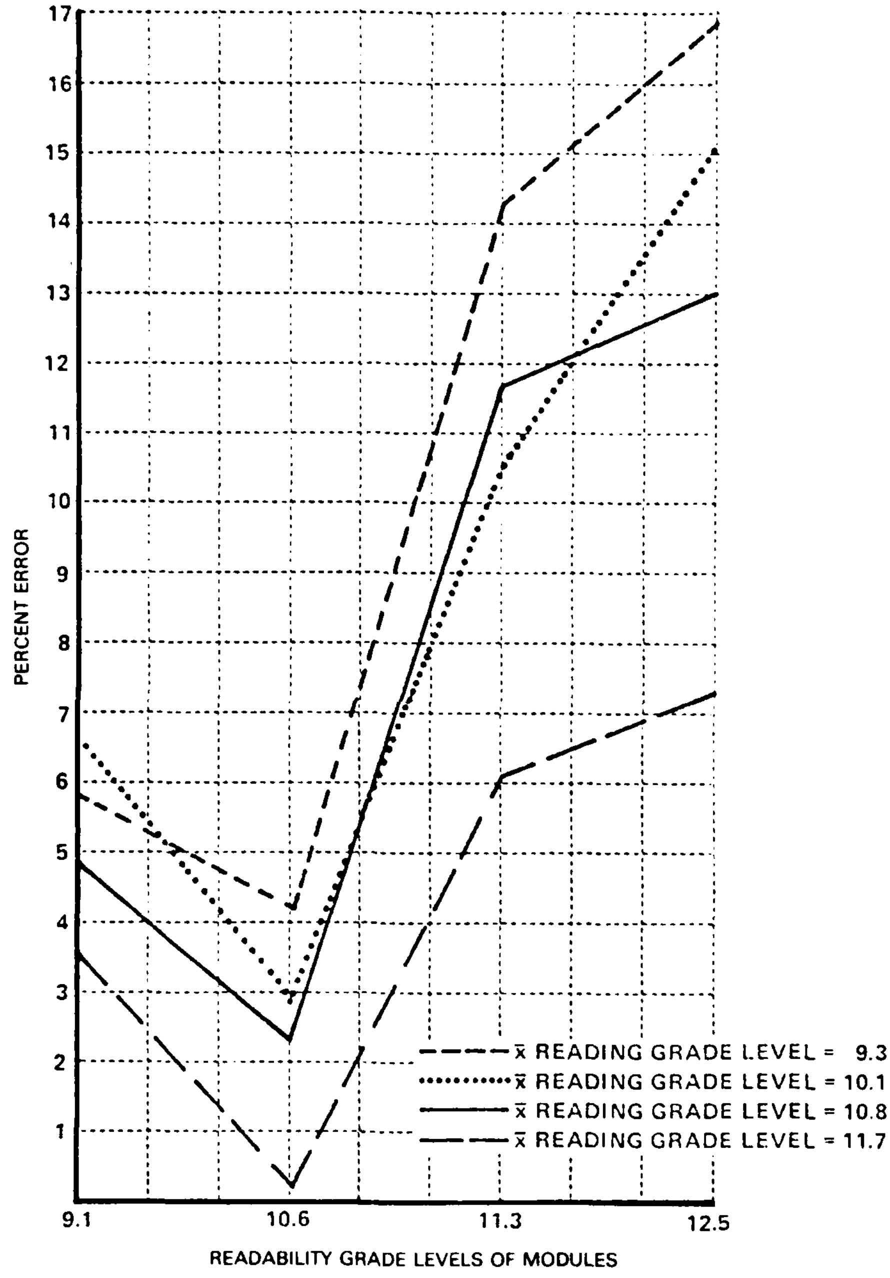


Figure 11. Comprehension test scores (percent error) for four reading grade levels of subjects as a function of modular readability.

Table 8

Mean Time-to-Criterion (Minutes Per Learning Objective) for Levels of Reading Ability and Readability

	RGL = 11.7	RGL = 10.8	RGL = 10.1	RGL = 9.3
Reading Ability	3.14	3.84	4.25	4.69
	MGL = 9.1	MGL = 10.6	MGL = 11.3	MGL = 12.5
Readability	1.94	2.13	4.95	6.91

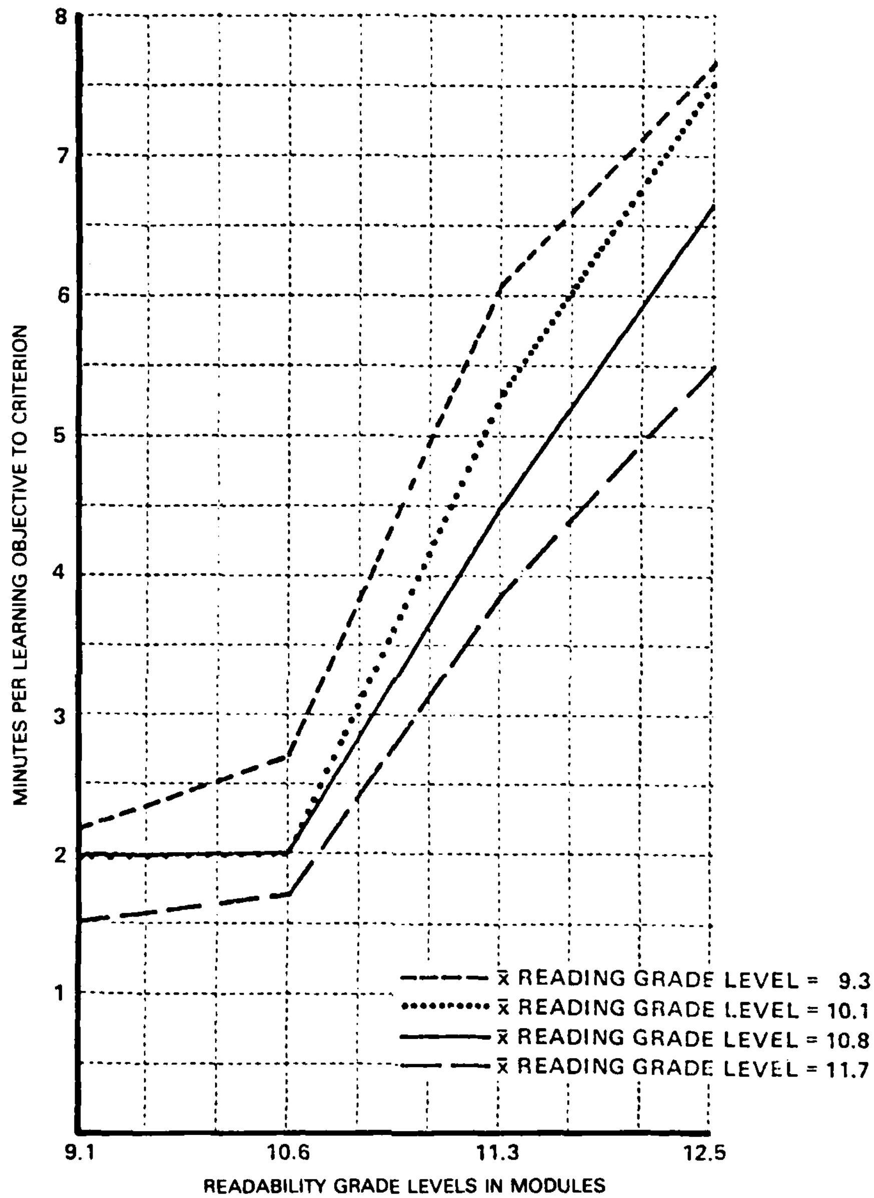


Figure 12. Learning time to criterion (minutes per learning objectives) for four reading grade levels of subjects as a function of modular readability.

Discussion

The results of this study further establish the validity of the recalculated version of the Flesch Reading Ease Formula in discriminating between passages of varying levels of difficulty. Using the measures of comprehension and learning time, this formula successfully "predicted" performance relationships as inferred from its grade level designations of modular programmed instructional materials. That is, student performance predictably declined as readability demands increased, especially for the lower ability readers. More importantly, these findings establish that learning time relates to readability in an equivalently systematic manner.

The two-fold purpose of this study involving validation of the NRIs and investigation of the relationship between readability and learning time appears to have been fulfilled. While the recalculated Flesch Reading Ease Formula was used as the primary indicator of readability throughout the study, the results should generalize to the recalculated Automated Readability Index (ARI) and the recalculated Fog Count as well. This, of course, is due to the criterion derivation of all three formulas from the same data base. Particularly important also is the extension of readability formulas to the programmed instructional domain.

Regarding the criterion of learning time, this study has demonstrated significant predictability for an additional performance measure for readability formulas. Further research is now indicated in two primary areas:

1. Validation research should be conducted using readability formulas in the rewrite of narrative passages to several grade levels for experimental test of the learning time criterion. Additionally, efforts should be made to control the effects of prior knowledge and learning styles. Then, multiple regression equations should be developed from a wide range of readability measures for eventual application in the operational setting.

2. Extensive investigation into the use of readability as a structural variable to complement individual difference variables in multiple regression research also appears to be warranted now.

MEASUREMENT OF TECHNICAL READING ABILITY AND THE DIFFICULTY OF TECHNICAL MATERIALS: A RESEARCH DIRECTION

Thomas E. Curran

Navy Personnel Research and Development Center
San Diego, California 92152

ABSTRACT

There is no question that literary skills among young people in our country are declining. Data relating to this decline, however, are based upon "general reading and writing ability." There is good reason to believe that, beyond such general abilities, there are attributes that are related to abilities in a person's chosen field. At issue here is the specific ability acquired by a person working in a technical area in which he acquires a vocabulary unique to that field that expands his literary capacity in that field.

This paper presents a research approach that might help to clarify the added talents achieved by a person in a technical field that are not indexed by recognized standardized tests. This is important because of the apparent mismatch between the readability formula scores for written materials that must be used by such persons and the "reading grade levels" at which their test scores place them. It is believed that they acquire literary abilities through association with their fields, which enable them to understand these materials despite the mismatch that appears (on paper) probable.

Introduction

There is a great anomoly in our country today. The United States of America boasts of the greatest technological society ever known to man and yet the intellectual capacities of the bulk of our young people appear to be on the decline. Precise statistics describing the waning skills of our high school graduates are difficult to acquire. The newspapers, however, are filled with indications that, if true, are somewhat alarming. United Press International (25 August 1977) reported, for example, that scores on the Scholastic Aptitude Test (SAT) for young people entering college in the Fall of 1977 were the lowest in the 51-year history of that test. The scores of students for the verbal portion of that test dropped two points from 1976 to 1977, and the average score has dropped 49 points since 1963. And one must remember that these are students who intended to enroll in college in the subsequent fall semester.

These statistics are disquieting, and yet would seem to represent an even brighter picture than that faced by the Armed Forces. It was reported (New York Times News Service, 16 November 1977) that more than 40 percent of the recruits in the volunteer armed forces now drop out during their first term of service. This figure has almost doubled since the early days of the Vietnam war. Literacy problems, of course, do not account for this entire figure; medical and disciplinary problems, poor performance (which is quite likely related to intellectual ability), etc. also play a part. But many of these "drop-outs" are due simply to the inability of the man to cope with the intellectual requirements of the Navy. Faced with these facts, one must wonder how it is that our increasingly sophisticated systems and equipments function as well as they do, when the skills of our operators and maintainers are at such seemingly low levels.

The armed services, of course, have their own "entry level" tests. The one in use today is the Armed Services Vocational Aptitude Battery (ASVAB), for which little longitudinal data are yet available. The ASVAB replaced the General Classification Test (GCT) and the comparisons here will use this earlier test. From a person's score on the GCT, a person's reading ability (which is the aptitude we are most concerned with here) can be predicted with a fairly high degree of accuracy (with a correlation coefficient of .73; see Duffy, 1977). Using the GCT score to make such predictions, Duffy determined that the median reading ability of recruits in San Diego was at the 10.5 level with 25 percent reading below the 8.7 grade level. He states:

> This can be compared to the reading difficulty of materials faced early in training (10.2 to 11.5 grade level) and to the difficulty of training school materials (average 14.0 grade level). These data indicate a sizable literacy gap, with 25 percent of the recruits reading five grade levels below the average difficulty of the training materials. (p. v)

The patterns of abilities for Navy recruits would seem to be following the trends in the general population. Looking just at GCT scores over the past years, however, the averages for entering recruits is remaining virtually the same, without the decline found in tests such as the SAT (Dr. E. Aiken, personal communication, July 1978). Regardless of the reasons behind either the general decline in scores on the SAT or the apparent lack of such decline in the GCT, the position to be taken here is that neither of these is important to us in examining the abilities of experienced Navy personnel. It is to be contended here that the abilities (reading, reasoning, etc.) that enable our experienced technical personnel to at least adequately carry out their jobs are not being tapped by, or are developed subsequent to, the administration of the GCT, ASVAB, SAT, etc.

The apparent differences between the difficulty levels of Navy Rate Training Manuals (RTMs) required to advance to petty officer 1st Class and chief petty officer (as measured by readability formulas) and what is known about personnel abilities will be the focus for a possible solution to this problem.

Background

The most comprehensive examination of the difficulty of Navy RTMs was reported by Biersner in a 1975 report for the Chief of Naval Education and Training Support (CNETS). This was in response to a CNETS request (7 June 1974) that "action be taken to develop and implement a plan to establish the reading grade levels of all rate training manuals . . . for which CNETS has management and publishing responsibility." As a result, the reading grade level (RGL) of 185 RTMs (of a total of 188) were determined using several of the most widely known formulas.

The Biersner study is considered to be a valuable and critical one when viewed in terms of the state-of-the-art in the field of readability measurement. Comments in that report itself, however, as well as those from a variety of sources at later dates, point out dangerous shortfalls of the work. As Biersner points out,

> . . . the relationship between RGLs (as determined by <u>any</u> of the reading formulas which are available) and reading comprehension or performance effectiveness is not well established, despite the apparent importance of reading to the development of most other skills. (p. 7)

He goes on to point out that the RGLs for the majority of the manuals appear to be at the lower college level (13th and 14th grades). When one considers this at face value in the light of what we know about the abilities of our entering recruits and of that age-group population in general, there seems to be cause for alarm. One reaction to these data came from a memorandum prepared by the Chief of Naval Personnel (CNP) for his assistant chief for Personnel, Planning, and Programming. In this memorandum, it was stated that the mean RGL for all the RTMs considered by Biersner is inflated by manuals for ratings "which are either extremely technical or those which have a language all their own." This statement is certainly true and puts another hole in the readability measurement "dike" without promising any mortar for plugging that hole in the near future.

Biersner himself hits upon the key issue here, when he says that:

> Most of the RTM material is not read by recruits, but is read instead by personnel who may have improved reading skills after leaving recruit training. The disparity between the readability of this written material and the reading skill levels of these experienced personnel may, therefore, not be as large as originally assumed. (1975, p. 28)

To examine this point more closely, let us examine the inflation of RGLs caused by the highly technical or unique languages of occupational groups. It is certainly true that the readability formula score is affected by these factors, because one determinant of that score in virtually all readability formulas is word length, and, of course, technical terms tend to be long. This is not the same thing, however, as saying that the readability (or difficulty) of these materials exceed the abilities of the personnel required to read them. To illustrate, the Disbursing Clerk 1st and Chief RTM may be easily comprehended by the experienced Disbursing Clerk (DK) despite the fact that it has the highest measured RGL (16.26) of any of the RTMs examined by Biersner. The "College senior" RGL does not necessarily mean that a person must score at the college level on a standardized reading test in order to comprehend it. The language peculiar to disbursing, however "foreign" to a person not in that field and however inflating to traditional readability formulas, may well be second nature to a DK2 or DK1.

A number of sources appear to have accepted the above reasoning, but so far have put forth no solution to the problem. Curran (1976) for example, points out that RGLs as determined by readability formulas have no one-to-one correspondence with "the ability of a person to profit from the written material." He states that "it may well be that the mere presence of a Navy recruit in the naval mileau may be sufficient for him to comprehend the necessary elements of the RTM required for advancement." Statements of this type do not go far enough. If this is true of a Navy recruit (that his abilities improve by virtue of his exposure to Navy life), it should be even more significant for those who have been in the service for 5, 10, or more years.

Quite simply, we have been talking about this problem for too long. It is clearly time that steps are taken to determine realistically whether or not, and if so, where and to what degree, a readability "gap" actually exists among our experienced personnel. Then and only then can positive steps be taken to reduce or eliminate such a gap. Perhaps it will be here at this workshop that some progress toward solution of this problem will occur.

Rationale: Past, Present, and Future

For the past 3 decades or more, attempts have been made to determine the reading difficulty of written materials for the typical intended reader of that material. Most of this work has involved the application of "readability formulas" to the material in question and obtaining an index of its difficulty. There is considerable disagreement as to the appropriateness of using such readability formulas in determining the difficulty of materials to be read by adults, even if their use for elementary school materials may be justified. The position taken here will avoid the extreme position that readability formulas are not applicable at all to military materials. Rather, it will be accepted that readability formulas are of <u>some</u> value, in the absence of more appropriate techniques, providing that they are used for materials to be read by persons similar in every possible way to the norm group upon which the formula was developed.

The formula to be examined most closely here is the revision to the Flesch "Reading Ease Formula" as reported by Kincaid, Fishburne, Rogers, and Chissom (1975). While the derivation of a formula using Navy written materials and Navy enlisted personnel was clearly needed, the question of the applicability of such a formula to advanced materials still remains. The norm group used by Kincaid et al., for this study were "predominately new enlistees with less than 6 months in the Navy." The bulk of the materials upon which these persons were tested were advancement manuals for petty officer 3rd class and above (of the 18 manuals used, six were for PO 1&C and seven for PO 3&2). The criteria of the difficulty of these passages were scores on a Cloze test, which required that examinees fill in the exact missing words of the passages. By virtue of the design of such tests, words such as "communique" (Journalist 1&C), "stratospheric" (Aerographer's Mate 1&C), and "contaminants" (Machinery Repairman 1&C) had to be supplied in order for the answer to be correct. One would expect an E-5 in those ratings to be much more likely to supply the correct word than would the novice. The true difficulty of the passages for the <u>intended</u> audience can be ascertained only by the responses of the more expert man. In other words, we are not at all interested in how difficult an RTM for 1&C (required for men with some 5-6 years' experience in the rating) is for a man who has been in the Navy for 6 months or less. It must be concluded here that using the Kincaid formula to measure the difficulty of written materials used by Navy recruits is of value; using this formula to measure the difficulty of more complex materials, however, is both incorrect and misleading.

Future Directions

1. Every effort should be made to make known exactly what a readability formula score does and does not mean. We must avoid having the uninitiated make hard and fast comparisons of "reading grade levels" of written technical material and tested reading ability of the experienced man who must use those materials.

2. Work must continue to determine the degree to which technical terminology inflates readability formula scores. Depending on the outcome of this work, clear guidelines must be established for the use of formulas in measuring the difficulty of technical documents and for the care which must be exercised in interpreting the resulting scores.

3. Instruments for the measurement of the "technical reading ability" of experienced personnel must be developed. Once such instruments are available, it is imperative that abilities determined by such instruments and the difficulty of the materials which such personnel must read can be compared in some meaningful fashion.

4. Data resulting from the above comparison must be made available in some usable fashion to those responsible for the writing and editing of technical materials to result in technical reading materials "matched to the man."

Editor's Note: Dr. Curran pointed out that his paper represented a research rationale for what he considered a major problem, with little or no guidance as to a methodology for solving that problem. He said that it represented, for him, an incomplete task with the "intellectual itch" associated with such lack of closure. The following discussion, occurring at several points in time during the course of the workshop, pertains to the general problems pointed out in that paper.

Dr. Curran: (Would you agree that the Kincaid formula is appropriate only) when we're talking about measuring the difficulty level of basic materials--the kinds of materials with which that formula was developed?

Dr. Kincaid: The basic subject population was made up of fairly inexperienced naval enlisted personnel, predominantly in "A" school. And you're right, I think, in your assumption that they hadn't learned the precise technical vocabulary for a particular specialty that you would expect a senior man on the job would have learned after 10 years. What the analogy would be--we all have our own particular areas of expertise. We find reading material in this area very, very easy. For example, I can skim articles in readability and pick out the pertinent points, but if I go to another technical skill, I can't (read it well) at all. I'm sure this is an experience familiar to us all.

(Different skill areas) have their own specialized vocabularies. And there is no doubt it becomes more of a factor with people who are more experienced and at higher levels of responsibility. There is no debate on that at all. And this particular formula, and others, are relatively general and are, in fact, most pertinent to relatively inexperienced enlisted personnel. But, on the other hand, if you get formulas that are directly pertinent to a series of technical areas--you might be able to narrow it to four or five--and you want to apply these in the form of standards, you have a very cumbersome situation.

Dr. Curran: It seems as though it would almost be subject-matter-unique, would it not?

Dr. Kincaid: Hopefully, you can take clusters of job specialties that have something in common, like electronics, and come up with a vocabulary (for each cluster). I'm going to be trying to do that in the next 6 months (in developing computer-stored technical word lists). So I certainly hope it can be done. But there is no doubt that if you're going to assess material for experienced specialties, a whole series of formulas would be required. Now these might very well be useful (for general readability measurement). But I question whether there might not be some misunderstanding if they were to become standards imposed in contracts. There is no doubt that they could be developed and would be useful otherwise.

Dr. Klare: I think there is a question also of how much more predictive they would be than existing formulas. For example, people have attempted to use the Dale-Chall with science terms (included in the word list) to make the formula more predictive, and it has not increased its predictiveness. This doesn't mean, of course, that it shouldn't be tried at the level of technical material. But what formulas do is operate on a kind of "basic core" of difficulty that is predictable by these index variables (word length and sentence length), and you just can't seem to go much higher than that. So, I doubt that you would increase your predictiveness very much (by trying to account for technical terms). I'd like to see it, however, because it needs to be done at that level, and I hope (Dr. Curran) will do it. But I personally would be rather surprised if you increased your predictiveness very much. Now you might get a better grade level index. In other words, you might be able to say: that's not really 16, it's actually 15. But, you could probably accomplish the same goal by simply renorming existing formulas--a more simple task than looking for a whole new set of variables.

Dr. Curran: It seems as though there are several directions we could go here. First, I think it would be of at least theoretical interest to renorm the Kincaid formula.

Dr. Klare: I do too, and I think that would provide more useful kinds of results than trying to add in more variables.

Dr. Curran: And I would clearly like to see the differences in Cloze scores between the sample used by Kincaid and experienced personnel (in those ratings from which Kincaid's materials were taken).

Dr. Kincaid: That raw data will made available to you.

Dr. Curran: I'll certainly take advantage of that offer.

Dr. Klare: Another question comes up. Are you interested in the difficulty level of the material other than the technical terms which (the experienced people) might know? If you are, you can . . . skip the technical terms and continue your syllable count on the nontechnical terms beyond, and get a syllable count to give you the index for all but (the technical terms). It would be interesting to see, if you tried that, if you indeed have a different kind of measure.

Dr. Kincaid: A first problem is to identify what you're calling "technical terms."

Dr. Klare: I think you could have judges do that. And then, if you haven't increased your predictiveness, it's a nonproblem, and you know you won't have to go through all that in the future. If it does make a big difference in the predictiveness of the measure, then you would have to go to work on the technical problem of (identifying technical terms in a way that would not require judges on each sample to be tested).

Dr. Kern: On some work that we've done, using a "modified Flesch formula," that sort of thing resulted in a drop of two to three grade levels.

Dr. Curran: A major consideration in this whole endeavor is whether or not we'll have to go through that procedure for every single technical rating. At best, it may be possible to cluster ratings so that, for example, all electronic ratings would use the same word list. That would at least reduce the number. But it almost certainly is going to be rating group specific.

Dr. Kincaid: Let me make one suggestion (for a possible research direction). Identify the "technical terms" in a field and then test knowledge of those terms as a function of experience levels.

Dr. Curran: I've been thinking along those same lines--of using as one possible instrument a "word recognition test" of some sort, using as many technical terms as I can extract from (written materials) in that particular rating.

Dr. Kincaid: That should give you some useful insight into their knowledge (about their ratings).

Dr. Curran: I would like to get your (Dr. Kincaid's) data (from recalculation of the Flesch formula). Then I would like to compare the performance of advanced levels (pay grades) from several ratings on the Cloze test, and on standardized reading tests to see if they are different. If they (the experienced personnel) do remarkably better on the Cloze test, then that means that the criterion, if you're using the formula to apply to (their materials), is wrong, as I see it.

Dr. Klare: I would suggest that you keep in mind two different possible goals: one, to increase the amount of nonerror variance, or the predictiveness of the formula, or two, whether you simply want to get a more accurate grade level (for specific ratings) than we now have. The latter, I think, is much easier to do than the former . . . and might help to overcome the problem of people thinking of grade level as "gospel."

SYNOPSIS OF AFHRL LITERACY RESEARCH PROJECTS

Sharon L. Slaughter

Air Force Human Resources Laboratory
Lowry Air Force Base, CO

ABSTRACT

This paper centers upon three ongoing research projects in the Air Force Human Resources Laboratory Literacy Research area and two new projects soon to be undertaken.

Briefly, the three ongoing projects deal with:

1. Making on-the-job reading materials more job-relevant.

2. Constructing materials with which a person can, on his own, improve his reading comprehension and memory skills.

3. Developing a criterion-referenced test to determine if written materials reach acceptable readability standards.

The two new projects deal with:

1. Identification of actual AF job literacy needs.

2. Measurement of readability of nonnarrative text.

The Air Force has defined its literacy problem as a discrepancy or "gap" between the reading requirements or demands of training and job printed materials and the reading skills possessed by the personnel who use these written materials. The general research thrust of the Air Force to reduce this discrepancy has been a two-pronged approach--one dealing with the simplification/modification of materials to reduce the reading demands of the printed matter, and the other to implement training programs aimed at increasing the literacy skills of the individual.

This paper is intended to provide information about on-going literacy research efforts at the Air Force Human Resources Laboratory (AFHRL). A brief discussion of three work units will be provided. Also, there will be a discussion of two literacy research efforts that will soon be undertaken at this organization.

Work Unit 1121-04-10, Development of Job-Relevant Reading Materials, was begun in January 1975. The objective of the effort was to develop job-related reading materials for potential introduction into Air Force reading improvement programs. The basic rationale for this effort was that the use of job-relevant materials would serve to increase personnel motivation to continue in voluntary reading improvement programs, and to improve job performance and promotion potential through the acquisition of prerequisite reading skills. This work unit was an outgrowth of Work Unit 1121-04-05, Development of a Methodology for Measuring Reading Skills and Requirements in Air Force Career

Ladders, which identified the literacy requirements of various career ladders and the reading levels of personnel within those ladders.

Materials from the 64XXX Career Field (Supply) were examined to generate: (1) a self-paced handbook of techniques for improving reading comprehension and memory, (2) a job reading task test for pre- and post-test evaluations, and (3) a collection and analysis of follow-up data for the Job-Oriented Reading Program (JORP). A complete report of the development of the prototype JORP is provided in AFHRL-TR-77-34. Milestones 1 and 2 (above) are covered in that technical report. An in-house report, titled, "Additional Data for Evaluation of the Job-Oriented Reading Program," is being completed. This report addresses the third milestone and provides discussion on the norming of the JORP Test, the validity of the JORP Test, and the performance of the student group compared to other reference groups. The estimated completion date of this report is January 1979.

Work Unit 1121-04-13, Operational Consequences of Literacy Gap, started in June 1977. The objective of this effort is to determine, in operationally definable terms, the consequences of literacy gaps of varying magnitudes on reading comprehension. "Literacy gap" refers to the problem that exists when the reading grade level (RGL) of personnel is discrepant from the RGL of the technical and training materials they must use.

This study, as originally planned, involved three independent variables at each of three levels: (1) Air Force personnel at the 6th, 8th, and 10th RGLs, (2) Air Force job-related materials written at readability gaps of 0, -2, and -4, and (3) reading times of 30, 45, and 60 minutes, with testing occurring after every 15 minutes of reading. Reading comprehension, as measured by a 52-item multiple-choice test, and reading preference, as measured by five questions, served as dependent variables.

The technical report outlining this research and providing recommendations for Air Force managers is being completed. The estimated delivery date for this product is April 1979.

Work Unit 1121-04-14, Understandable Publications Validation Test, began in August 1977. The objective of this effort is to determine if publications, rewritten in accordance with HQ USAF Operating Instruction 5-1, are comprehensible to intended users.

A criterion-referenced evaluation approach, rather than a classical experimental (i.e., comparing old with new publications), was used. The device for determining the comprehensibility of Air Force publications was the Cloze procedure. The major milestones of this effort included: (1) review and selection of publications for field study, (2) development of Cloze tests, (3) data collection and analysis, and (4) technical report preparation.

Publications were reviewed and checked for reading difficulty level. The FORCAST reading difficulty level formula was used as the quality control check. Of the approximately 45 regulations reviewed, seven regulations were selected for field study purposes. Selection of regulations was based upon how closely the reading grade level (RGL) of the publication matched the RGL of the user audience, and the HQ USAF requirement to take a look at specified career fields.

Thirty-two Cloze tests were developed from the seven regulations. One vocabulary test was developed for each career field tested (seven). Subjects (N = 1359) were tested across USAF commands, skill levels, and duty station locations.

Data analysis has been completed and the technical report, outlining the specifics of this research, is being written. The estimated completion date of this report is February 1979.

The following is a discussion of literacy research efforts that will be initiated by FY79:

1. Work Unit 1121-04-15, Functional Literacy Task Inventory. The objective of this effort is to identify the actual literacy (reading and oral) behaviors that are demanded by AF jobs. This work will constitute the initial phase of a program that will, ultimately, result in a definition of functional literacy in the AF. This effort will require:

a. The conduct of a literature review.

b. The consideration of alternative occupational and reading task analysis methodologies and survey formats.

c. The development of a task inventory to identify literacy tasks that are required in the performance of AF jobs.

d. A field tryout of the inventory.

e. Documentation of the research.

The estimated start date of this 19-month effort is December 1978.

2. A purchase request (PR) package is being prepared for the Readability of Tests study. The objective of the effort is to develop methods for assessing the readability/comprehensibility of nonnarrative (test and checklist) materials. This effort will require accomplishment of the following technical tasks/objectives over an estimated 16-month period:

a. A review of readability/comprehensibility literature to identify prior approaches and text difficulty factors that may be applicable to the problem of measuring the reading difficulty of nonnarrative prose.

b. Identification of the structural, grammatical, psycholinguistic, or textual factors impacting the readability of test and checklist type items, and approaches for separating reading difficulty from item content difficulty.

c. Quantification of these factors and specification of their interrelationships in contributing to the reading difficulty of nonnarrative materials.

d. Development and validation of a methodology for predicting and assessing the readability of nonnarrative prose.

e. Demonstration of the methodology with AF materials, and recommendations for the appropriate implementation of the methodology.

f. Proceduralization of the methodology into a handbook suitable for use by AF test developers and technical writers.

The estimated start date of this project is March 1979.

Often projects scheduled for beginning in fiscal year 1979 include the following topics:

1. Identification of new index variables that might apply.
2. Quantification of such variables.
3. Demonstration of the methodology with USAF materials.

AN ATTEMPT TO COMMUNICATE WHILE TRAINING AND ASSESSING WITHOUT SERENDIPITY

Walter W. Birdsall

Associate Deputy, Training Program
and Examination Department

Naval Education and Training
Program Development Center, Ellyson
Pensacola, FL 32509

ABSTRACT

The following paper represents a narrative history of one who is involved in the day-to-day problems of maintaining an appropriate difficulty level for Navy rate training manuals and advancement examinations. It chronicles some of the difficulties involved in this enormous undertaking and presents some of the possible solutions to the problems involved.

- -

My remarks are personal and do not reflect attitudes, policies, or postures of either the Naval Education and Training Program Development Center specifically or the United States Navy generally.

As this is a workshop on readability (whatever that means), I must regard myself as an interloper, for my interest and concern are with communication--particularly in training and assessment.

Many attempts have been and are being made to ensure that the Navy's advancement examinations and rate training manuals communicate and perform their intended functions. Each of these two products have built-in problems but also share certain characteristics. Therefore, each must be addressed, initially, separately.

The Navy advancement examination is a management tool used to rank-order an already qualified population in terms of knowledges required of a given occupation, at a given level.

A Navy occupation is very broad-based by necessity. A ship is basically self-contained and can accommodate just so many people. Therefore, these people must be able to perform all functions required for the ship to operate, fulfill its mission, and service its own personnel. An example of such an occupation is the hospital corpsmen in which there are about 40 subspecialities ranging from ward attendant and dietician to sanitation engineer and pharmacist, yet a corpsman on independent duty must be able to function in all areas of these individual spheres. To assess such a diversified population, regardless of assignment and geographical location, requires a precision of language that is not as pronounced as in text writing. The words used must communicate the same thing to all but also also must be in the vocabulary experience of all.

In 1955 there was concern that the composition of the then Steward rating was being penalized severely on a test-score basis on the attribute ostensibly being measured because of linguistic and reading ability handicaps. At that time only two nonverbal group tests were available--the Army Beta and the Semantic tests of intelligence in which the directions are given completely in pantomime. Other purported nonverbal tests were merely partial verbal. A pictorial examination was developed and administered. The results were inconclusive and because of the great expense in producing such a test, continued experimentation with the format seemed unwarranted.

As the Filipino composition of this same rating group has steadily increased, another attempt was made in the mid 60s; that is, an examination was written in Tagal. But because of our extreme naivete as to the dialect problems of the language--let alone that there were 42 identifiable and distinct dialects that took on the aspect of a patois, the experiment must be considered the ne plus ultra for design failure.

The foregoing were specific attempts to address specific language problems. At the same time, all examinations were being analyzed in terms of reading difficulty. In the late 40s, Flesch's "Readability Yardstick" thermofaxed for all hands was de rigueur in the testing bag of tricks. This was succeeded quickly by Dale-Chall, "Thorndike's Word List," and most recently, Kincaid's "Average Grade Level." Concomitantly, the Thorndike-Barnhart Junior Dictionary constituted the nonprofessional word parameters of the examinations.

In the 30 years that the advancement examinations have been developed, over 3 million questions have been generated and individual item analysis gathered. Sporadically, these analyses have been correlated with the "reading level" of the question. This was done in all sorts of configurations (i.e., within rate, within rating, across rates, across ratings, all types of matches, and mismatches). All results were inconclusive. The data did not produce evidence that a reading formula be utilized in the development of a question.

The rate training manuals are designed to provide an overview of an occupation at a given pay grade. They are general in nature and are not designed as "how to" books for specific equipment but, rather, to acquaint the student with a prototype. Also, these manuals compile and digest material from various technical sources so that this material is more accessible and meaningful. An inherent problem in the design of a training manual is determining whether the manual should be at the level of the reader-learner or should force and/or assist the reader to a higher level. An attempt is made to strike a balance between these two.

These manuals are used in Class "A" school curricula and are the primary source of rating information for personnel who do not attend a Class "A" school. It is rather interesting that entrance to Class "A" school is governed in part by GCT score. The gearing is that the Navy sends those to school who can most benefit the Navy, not necessarily those who could most benefit from Class "A" school. The fallout of this is that the less gifted derive their rating knowledges from the manuals basically on their own and the more gifted are assisted by schoolhouse training. While this concept seems reverse of current educational and societal vogues, it is consistent and necessary in the military establishment. This policy forces the rate training manual to communicate to the less gifted of the population rather than to the general population.

Although the procedures for an established reading level for the advancement examination appear vague, they are more stringent than those that have been applied to the manuals over the same period of time. To cite from two directives concerning reading levels,

> . . . in writing a text for enlisted men, assume that the student is at high school level; and (2) in writing a text for officers, assume college training on the part of the reader. These rules are general

and should be modified in the light of your experience in the subject and your knowledge of trainees in the field.

You may find the following helpful in determining the reading level of the material you write.

Reading Material	Reading Level
Pulp (westerns, True Story)	
Slick (Saturday Evening Post, Collier's, Ladies Home Journal) . . .	6th-8th grade
Digests (Reader's Digest, Time) . . .	8th-10th grade
Quality (Harper's, New Yorker, Business Week) . . .	11th-12th grade
Scientific (Professional papers) . . .	College graduate (above 16th grade)

The conclusions you can draw are rather obvious. Widely popular reading material in this country goes no higher than the 10th grade. Only a relatively small portion of the population feels at ease with more difficult prose.

There are so-called "readability formulas" that may help you to determine the reading level of your manuscript. A readability formula is a statistical tool and should be treated as such. No formula is an infallible measure of reading difficulty. This is especially true when applied to technical writing because more words will always build up the word count. A low count does not guarantee clear meaning; a high one does not always create reading difficulty. No readability formula should be permitted to take the place of sound judgment.

When reading formulae were strictly applied, the result was short, sing-song, monotonous sentence structure that became a mental cant similar to the clickity clack of a train on a railroad track.

In 1974, the Chief of Naval Education and Training Support directed that the reading levels of all rate training manuals be determined. This was precipitated by the much publicized report of Carver who stated that 20 manuals studied with his formula yielded a 14th grade reading level. A very structured research was conducted and reported in <u>CNETs Report 2-75</u> (Biersner, 1975).

There were four parts of the research that were not included in the report. In addition to all of the Navy training manuals that were in print, reading levels were also ascertained on all the text books used in the 8th grade of all Escambia County schools. These paralleled the Navy's books in that they ranged from the 6th to college graduate level with an average level of <u>13th</u> grade.

A second part of the study was the reading level of an entire chapter of 16 pages in the Basic Military Requirements Manual. It was not sampled, but divided into 100-word segments and a level established for each segment. Then the same chapter was redone starting at the 50th word and establishing new segments. The first 50 words were added to the last segment. The surprise was that the first count resulted in an 8th grade level; and the second count, a 10th grade level. This difference from merely slipping the starting point 50 words.

The third unreported aspect was that reading levels were made on random samples of 8th grade reading comprehension tests. These too yielded a range of reading scores.

The last unreported element of the study was actually a question. When a reading level is reported, what is meant? Is the same meaning to be derived when it is stated that a 14-year old--who is in the 8th grade--is reading material that has an 8th grade reading level as when a 20-year old--who is a high school graduate--can only read at the 8th grade level? This question and any permutation of these six variables should yield interesting results.

After the study was concluded and disseminated, it was interesting that these reading levels were being interpreted as comprehension levels. In fact, it seems that this has been a widely-held concept that reading level and comprehension level are tantamount and synonymous.

An interesting aside. We are currently developing a new edition for Basic Electronics and Electricity. This effort requires a great deal of research including analyzing existing commercial texts. Many of these texts report a reading level for which they are designed. In reading these texts that were reported as being at the 8th and 9th grade level, they seemed to be at levels equal to other texts that were pegged at a much higher level. We then did Flesch on these reported 8th and 9th grade level texts and found they exceeded the 13th level. I contacted three different publishers to inquire how they had established the reading levels. Two publishers stated that they had tried various formulae and reported out the findings of the formula producing the lowest grade level. The third publisher stated that he would be glad to discuss it with me personally were I ever in the area (1000 miles away), but the policy of his publishing house was not to casually discuss these matters over the phone.

As our goal is to communicate, we are still trying. Two different types of approaches are now underway. The Navy has over 200 Navy Junior Reserve Officer Training Corps Units in high schools throughout the United States. Some of these units are ghetto-based. The students in these units will act as reviewers to our new edition of Basic Military Requirements. It is hoped that the open comments from these students will assist us in designing material that not only communicates to them, but also appeals to them.

A second approach is a little more radical. A manual is being developed without educational or psychological expertise. A journalist-public relations type is going over factual material to ascertain what material lends itself to pictorial format. Then an illustrator will render this into some sort of graphic presentation. The result will be, hopefully, a picture book with words used as bridges rather than a book with pictures.

In all of our products we want to and must communicate with and train and/or assess the user. Any technique that will assist our doing that will be welcomed.

I appreciate the invitation to participate in this workshop. Such an endeavor as this meeting should have the result of research in action.

BUYING READABILITY—SPECIFICATIONS AND STANDARDS

Michael Mecherikoff

EG&G Washington Analytical Services Center, Inc.
500 Esplanade Drive, Suite 240
Oxnard, California 93030

ABSTRACT

Buying tech manual usability, including readability, is difficult because very few contractual requirements can be written that are specific enough for objective evaluation of contractor compliance and whose specificity is derived from known impact on user performance. Systematic research, generally of a rather dull sort, can furnish much of the needed information. Such hard information is needed to support a wide variety of categories of requirements. The efforts to refine traditional RGL formulas seem disproportionately large, and foster the continued misuse of an index that is of doubtful applicability and that, outside of a small research community, may be almost universally misinterpreted.

Introduction

Three years ago, under contract with the Naval Sea Systems Command Fleet Support Directorate, Technical Publications Branch, my company was tasked to develop a military specification to control readability characteristics of NAVSEA technical manuals. What was wanted was not a guide or handbook with advice and suggestions, but, rather, a document which could be invoked in contracts and whose requirements would be strictly complied with by tech manual preparers.

Two years ago I inherited the project, and by definition instantly became an expert in readability. My task at the time was to review the current draft specification for internal consistency and relationships to other tech manual specifications, to verify that the requirements were in line with certain source documents, and to ensure that the requirements were stated in proper contractual language--that is, language such that contractor compliance with the requirements could be unequivocally evaluated. The document would then go to various agencies within NAVSEA for review and comment, the comments would be evaluated and incorporated, and the draft would go to a NAVSEA specification review board for approval for use within NAVSEA. During the past 2 years, the document has been redrafted as a military standard rather than a military specification and is now up for approval by the specification review board. It has not yet been used to procure a tech manual, so I can't report on its effectiveness.

What I <u>can</u> report on are some of the headaches I had trying to produce a set of contractually binding requirements out of our generally skimpy fund of hard knowledge, plus readability formulas of dubious applicability, plus a body of suggestions and advice on technical writing style (some of it very <u>good</u> advice), plus some prior attempts at legislating readability.

As if this wasn't trouble enough, there were also requirements intended to ensure the comprehensibility of illustrations. Tom Curran will report on an interesting excursion we took into that area, so I'm going to limit my remarks to textual comprehensibility.

I'm interested only secondarily in describing the specific formulations that appear in the current draft of the military standard, because these are subject to change. What I <u>do</u>

want to mention are some of the options--some of the alternative ways of formulating the requirements.

Categories of Readability Requirements

Working with these requirements, various distinguishable categories of requirements emerged, reflecting the features of technical manuals and technical writing that we currently think might be important in influencing comprehensibility. These categories probably don't comprise a theoretically important taxonomy, but are just--for me--a convenient list of things we think we want to get control over. For example, we want to get control over vocabulary, syntax, content (which includes placement, organization, and relevance of material), cues to content (by which I mean such user aids as paragraph headings and topic sentences), and something called "level," which of course is not independent of the previous items, but which is often talked about as if it were a manipulable entity.

You've noticed, I'm sure, that I've drifted beyond a restricted definition of readability and am talking somewhere between the broader concept of comprehensibility and the even broader concept of usability. I've done that deliberately, because I'm convinced that, when we're struggling with a readability concept for technical writing, we have to keep it well anchored in the context of usability. One of the problems I have with the Reading Grade Level approach is that it tries to provide a measurement of considerable purity, uncontaminated by considerations of how the reading is done and for what purposes.

While we're still out here in the broader concept of usability, let me add a few more items to my list of features to control. We would like to be able to specify control over the overall organization of the document, the graphic illustrations (tables, charts, diagrams, pictures), the relationships between graphic presentations and text, and the facilities for accessing information (such as indexes, cross references, and the like).

We know--or at least strongly believe--that in all these areas there are right ways of presenting technical information and wrong ways, in terms of effectively supporting the performance of the technician who needs to use the information to do his job right. Our dual problem is, first, to find out with some degree of certainty what these correct ways are (of course as a function of personnel characteristics, working environment, the nature of the hardware, and the nature of the job tasks), and second, to formulate our knowledge in language that will contractually obligate preparers of technical data to do the right things.

Control of Vocabulary, Syntax, and Level

Returning for a few moments to a more restricted concept of readability, let me give some sometimes frustrating examples of what has been attempted, or is being attempted, in the control of vocabulary, syntax, and level.

Vocabulary

Control over vocabulary, including acronyms and abbreviations, takes several forms. One of these is to specify a maximum for the average word length, under the generally accepted assumption that, if words are kept short, they will also tend to be simple. The approach is reasonably objective, whether word length is measured by number of syllables or number of letters, and can be successfully automated. There is evidence that short words are in general easier to recognize and identify correctly, especially for poorer readers, quite apart from simplicity of meaning. That is, the rate of misreading the words is reduced if the words are short.

A significant problem is where to set the maximum. At present our understanding of average word length comes entirely from its appearance in Reading Grade Level formulas, but the RGL concept itself has serious difficulties when applied to adults and to technical writing. In short, we really have no idea how to write a sensible requirement involving average word length.

A few minor problems also intrude. What to do with mandatory long words? How do you score acronyms? Abbreviations? Numbers? Hyphenated words? Conventions for these can be adopted, but different authors have adopted different conventions.

A popular way of controlling vocabulary is by providing lists of preferred words. These lists have, to date, usually contained action verbs for directing the behavior of the technician. I was directed to include such a list in my draft, and the draft I was working from also had one. So I took a close look at the list. The first thing I noticed was that every verb was defined, even words like "cut" and "stop," which was defined as "to cease." I had a lot of trouble imagining a situation where a technical writer or editor would be unclear about the meaning of "stop" but where his confusion would vanish immediately when he consulted the definition in the verb list. An example of the use of each verb in a sentence was also given, even for the verbs which are understood pretty accurately by 3-year olds! A clear case of a list-maker's neurotic compulsiveness to run wild!

The third thing I noticed about the list was that words like "help" and "assist" were both there, and no preference was indicated. Are they both preferred? If so, what are they preferred to?

I had noticed a similar-looking list in a guide recently published by the Naval Air Systems Command, so I took a close look at that one. It turned out to be almost identical to ours, including the silly definitions, except for one important difference. The synonyms were gone. One synonym had been selected for inclusion and the rest left out. That seemed to make more sense, but from the standpoint of the user of the list, something was still missing. Suppose the technical writer wants to say something like, "Determine the length of the rod," but he decides to check the verb list. "Determine" is not there. What should he conclude? Not much. What he needs is a list that contains the word "Determine" with a notation that says "try using the word 'measure' instead." Interestingly, there is such a list in an Air Force spec, in which the synonyms are ranked according to their desirability in most cases. This list was apparently the source of the other two. The ridiculous definitions and examples were retained in the two derivative lists, while the really critical information--the preference value--was removed, apparently as a simplificaiton. In one case, the ranking was removed, leaving all the words with indiscriminate preference value. In the other case, only the first ranked word was included.

I've gone into this story in some detail because it illustrates something I've noticed about other lists provided for similar purposes. The lists are prepared carelessly, at least in the sense that there is little concern for how the list will be used by the intended user. A recently published military handbook related to the Army's very promising Integrated Technical Documentation and Training (ITDT) program, in addition to a verb list, also provides a list of familiar words which apparently purports to contain all the words, except technical nomenclature, that a technical writer would ever need to use. There are some strange entries such as "sunshine," "jockey," and "wee," but "submarine" is missing, as is "carburetor," and I would hesitate to classify these as technical nomenclature. Another interesting curiosity is that some of these lists are headed by a requirement that says something like, "The following words shall be used in technical manuals," which, taken literally, means that every technical manual must contain all the words! Now I realize I'm getting picky at this point, but all of this indicates that at least some readability requirements are being laid down by people who aren't paying close attention to what they're saying or doing, and we certainly don't need that.

Another sort of vocabulary control involves lists of words or expressions that are prohibited or at least discouraged. These lists tend to be short and to contain examples. I haven't seen one that attempts to be exhaustive. The aim is to reduce jargon and what is called "elaborate" or "pretentious" language. I have disagreed with various specific entries in such lists, but since this is a minor form of control, it's not worth pursuing here.

Sentence Syntax

Sentence syntax is also controlled in more than one way. The most quantitative approach is to control length, under the theory that it is difficult to generate a truly convoluted sentence of 15 words or less. I happen to like that theory, so NAVSEA's proposed standard has some requirements about average and maximum sentence length. Again we're guessing about what these values should be, but I feel a little more comfortable with these requirements than I do with those governing word length. The limits should probably vary depending on the purposes of the sentences. For example, descriptive material can probably stand longer sentences than lists of procedures. Some very practical, simple, and dull research would probably give us some idea of how to write such requirements a bit more intelligently.

Another approach to "syntax complexity problem alleviation" is to avoid stringing out a series of nouns as if they were adjectives! This is a strong tendency in military writing because writing becomes compact. For the same reason it puts a big load on the reader to find the end of the string and decode the meaning backwards. My opinion is that prepositional phrases and other devices are available in the language to solve just that problem and should be used. It is an opinion, however. When it becomes a requirement, it should be backed up with some evidence that certain kinds of readers can't handle such constructions, if that is indeed the case.

Discouraging the use of subordinate clauses should foster reading ease if sentences are long. The NAVSEA standard requires that sentences of more than 20 words that have clauses shall be broken up into simple sentences if possible. I don't have any idea if that requirement will do any good. Perhaps it's too clumsy an approach. Long ago I thought that computer scanning of text would soon result in useful quantification of syntactic complexity. I don't know exactly why that hasn't happened, but the results of various attempts through the years have been discouraging. It occurs to me that perhaps the attempts themselves were too complex. Generally there was more involved than simply flagging sentences according to certain moderately gross criteria, so that a writer or editor could take another look. A lot of text these days is in digital form at one time or another. There are computer programs that crank out Reading Grade Levels and other information based on a variety of methods. It's difficult for me to believe that our rather modest needs by way of syntactic analysis can't be handled almost as easily.

The sledgehammer approach to syntactic simplicity is to insist that sentences be framed as subject-verb-direct object, in that order, with modifiers as close as possible to the word modified. I used that one. Interestingly, the Air Force spec that did such a nice job on ranking verb preferences badly botched their syntax requirement, largely because they totally misinterpreted what an indirect object is. Unfortunately I've seen the same formulation picked up in other documents, including the predecessor draft that evolved into the present proposed standard. People believe and copy the wrong stuff as readily as the right, which emphasizes the need for good evidence and good methodology to underpin requirements.

Level

"Level" is one of those terms by which we sometimes kid ourselves into thinking we know what we're talking about. Let me illustrate with a paragraph from the NAVSEA proposed standard, along with a detailed comment from one of the NAVSEA codes, and my response to the comment.

The requirement reads:

> 4.4.1 Vocabulary. The simplest, most familiar, and most concrete words which accurately convey the intended meaning shall be used. Short words and words typically learned early in life shall be preferred. Use of highly technical terms shall be limited to those circumstances where simpler terms would not accurately convey the intended meaning.

This obviously falls short of perfect objectivity, but note the alternative suggested by one of NAVSEA's reviewers:

> Use of the phrase "words typically learned early in life" only confuses the issue. How early in life? How typically learned?
>
> In contrast, MIL-M-15071G, a widely used Navy TM specification states in paragraph 3.3: "Level of writing. The level of writing and development of text for types I, II, IIS, and III manuals shall be in accordance with MIL-M-38784 and the following:
>
> a. As a general guide, the level of writing should be for a high school graduate having specialized training as a technician in Navy training courses.
>
> b. Summary portions in chapter 1 shall be written to the level of command and supervisory personnel.
>
> c. Operating instructions shall be written to the level of an operator having previous experience in the operation of similar or related equipment.
>
> d. The level of writing for other portions of the manual shall be to that of a technician (Navy Technician Third Class) having previous maintenance experience with similar or related equipment.
>
> e. Type IIX manuals shall be written to the level of a graduate engineer familiar with the type of equipment involved."

My response to this alternative was as follows:

> The requirements which are quoted from MIL-M-15071G give the appearance of greater precision, but in fact are impossible to apply. There is little precise data on the reading abilities of the types of personnel mentioned, and no accepted procedures for guaranteeing that written material matches the

> requirements. For instance, "Operating instructions shall be written to the level of an operator. . ." is almost worthless in terms of specific guidance. Although the criteria in the proposed standard are still not precise, it is felt that writers, editors, and inspectors of technical manuals have better intuitive knowledge and can reach better argument with respect to "words learned early in life" than they can by trying to imagine the reading skills of, for instance, a hypothetical high school graduate or hypothetical command and supervisory personnel.

Obviously what we need here is some quantification, and unfortunately it's the Reading Grade Level (RGL) that provides the most common vehicle for achieving that objective. There are lots of things wrong with the association between RGL and technical manuals. Here are some of the main ones, in my opinion.

It's often said that the problem with RGL is that the formulas only measure a couple of factors and miss other crucial variables related to content and style. That's true, but the problem is worse than that. RGL is invariably a linear function of the word difficulty and sentence length measures that comprise it. That allows one to be traded off against the other. In other words, to maintain an RGL of 9, you could use sentences averaging 25 words long providing the words were short enough. Now, of course, writing isn't done that way, but the point is that such a composite is not really what you want in a technical manual. I used a maximum for average word length to help keep words simple, and one for average sentence length to keep the sentences simple, but did <u>not</u> combine them into a requirement for a maximum RGL. I think it's a mistake to encourage thinking in terms of this tradeoff, or even to provide the opportunity for it. People who don't know better will produce the kind of requirement we find in MIL-M-29355, which requires writing to an RGL of 7. The table reproduced here (Figure 13) is a precise recipe for performing such an illegitimate tradeoff.

The desire to have such a requirement stems from the fact that the term RGL is used to describe characteristics of both the reader and the material read. A 9th grader can be measured and be found to be statistically average in overall, composite reading skill, whatever <u>that</u> is. Reading material can be statistically evaluated and declared approximately suitable for average 9th graders. And within the limits of both measuring techniques, the match will work. An adult assigned an RGL of 9 by a test probably does <u>not</u> have the same pattern of competence within his overall, composite reading skill as the 9th grader has. The match probably doesn't work so well. But the Navy and the other services have never <u>intended</u> to match readers with an RGL of 9 to reading material with an RGL of 9. Worse, what they are matching is a large group whose <u>average</u> skill is represented by an RGL of 9, half of whose members are to some degree below an RGL of 9. This spread within the group is rarely talked about or apparently seriously recognized, except by a few of us scientists who are "in the know."

The concept of RGL itself is of course statistical. Some proportion, not 100 percent of 3rd graders, 4th graders, and so on, have read material and have taken tests on which they received scores, ordinarily not 100 percent reflecting their comprehension. In the military situation, the <u>ideal</u> case is for 100 percent of the users of a technical manual to comprehend 100 percent of the material so they can act 100 percent appropriately on it. We're not talking here about a developmental process of learning to read, where the population is children distributed according to a bell-shaped curve. We're talking in a sense about a transmitter and a receiver and some information, and we want that information to get from one to the other with very little degradation. That's perhaps a more accurate model for us than the school children. But the urge to match up those

Syllables per Word	Sentence Length (Words per Sentence)																				
	20	19.5	19	18.5	18	17.5	17	16.5	16	15.5	15	14.5	14	13.5	13	12.5	12	11.5	11	10.5	10
1.60	M11	M11	M11	L11	L11	H10	H10	H10	H10	H10	M10	M10	M10	--	--	--	--	--	--	--	--
1.57	L11	H10	H10	H10	H10	M10	M10	L10	L10	L10	L10	L10	H9	H9	H9	H9	--	--	--	--	--
1.55	M10	M10	M10	L10	L10	L10	L10	H9	H9	H9	M9	M9	L9	L9	L9	H8	H8	--	--	--	--
1.52	L10	H9	H9	H9	H9	M9	M9	M9	L9	L9	L9	H8	H8	M8	M8	L8	L8	L8	H7	--	--
1.50	H9	H9	H9	M9	M9	L9	L9	L9	H8	H8	H8	M8	M8	L8	L8	H7	H7	H7	M7	M7	M7
1.47	M9	L9	L9	L9	L8	H8	H8	H8	M8	L8	L8	L8	H7	H7	H7	M7	M7	M7	L7	L7	L7
1.45	L9	L9	L9	H8	H8	M8	M8	L8	L8	L8	H7	H7	H7	M7	M7	M7	L7	L7	L7	H6	H6
1.42	H8	H8	M8	M8	L8	L8	L8	H7	H7	H7	M7	M7	M7	L7	L7	L7	H6	H6	H6	M6	L6
1.40	M8	M8	L8	L8	L8	H7	H7	H7	M7	M7	M7	L7	L7	H6	H6	H6	H6	H6	M6	L6	H5
1.37	L8	H7	H7	H7	H7	M7	M7	M7	L7	L7	L7	L7	H6	H6	M6	M6	M6	L6	L6	M5	L5
1.35	H7	H7	H7	H7	M7	M7	L7	L7	L7	L7	H6	H6	H6	M6	M6	L6	L6	H5	H5	M5	H4
1.32	M7	M7	M7	M7	L7	L7	L7	L7	H6	H6	M6	M6	M6	L6	L6	L6	H5	M5	L5	H4	M4
1.30	M7	M7	L7	L7	L7	L7	L7	H6	H6	H6	M6	M6	L6	L6	H5	H5	M5	L5	H4	M4	L4
1.27	M7	L7	L7	L7	L7	H6	H6	M6	L6	M6	M6	L6	L6	H5	M5	L5	H4	H4	M4	L4	H3
1.25	L7	L7	L7	H6	H6	H6	M6	M6	L6	L6	L6	L6	H5	M5	L5	H4	M4	M4	L4	H3	M3
1.22	H6	H6	H6	H6	M6	M6	M6	L6	L6	L6	H5	M5	L5	H4	M4	M4	L4	L4	H3	M3	M3
1.20	H6	H6	M6	M6	M6	L6	L6	L6	H5	H5	M5	L5	H4	M4	L4	L4	H3	H3	M3	L3	L3
1.17	H6	M6	M6	L6	L6	L6	H5	H5	M5	M5	L5	H4	M4	L4	H3	H3	M3	M3	L3	L3	H2
1.15	M6	M6	L6	L6	L6	H5	H5	M5	L5	L5	H4	M4	L4	H3	H3	M3	L3	L3	H2	H2	M2
1.12	M6	L6	L6	L6	H5	M5	M5	L5	H4	M4	L4	L4	H3	M3	M3	L3	L3	H2	H2	M2	M2
1.10	L6	L6	H5	H5	M5	L5	L5	H4	M4	L4	H3	H3	M3	M3	L3	L3	H2	H2	M2	M2	L2
1.07	L6	H5	M5	L5	L5	H4	H4	L4	H3	H3	M3	M3	L3	L3	H2	H2	M2	M2	L2	L2	H1
1.05	--	--			--						--										--

Note. Number equates to grade level.
Letter to Grade Range
L--Low Grade
M--Medium Grade
H--High Grade
(e.g., M7 is medium 7 Grade Level).

Figure 13. Readability Scale. Table from MIL-M-29355 (MC) showing tradeoff between word length and sentence length.

RGLs is irresistable--almost as irresistable as the urge to modify, simplify, and renorm the old formulas! What I think we need is a new start from a different place. Leave the developmental processes to the educators and the kids, and develop some new indices of information transfer efficiency and relate them to the variables that enhance or degrade it.

Tom Curran and I were thrown back to this sort of new start last year when we were trying to figure out how to study comprehensibility of graphic materials. They don't teach that in school, and there are no RGLs, so we're back to a more fundamental level of questioning, thinking, and formulating methodology. What we've come up with so far is not very clever or sophisticated; it will take a lot of effort before it looks anything but simple-minded. But I think we have to get back to such a level in the reading area before we can get out of the trap which the RGL concept has become.

Conclusion

Buying tech manual usability, including readability, is difficult because very few contractual requirements can be written that are specific enough for objective evaluation of contractor compliance and whose specificity is derived from known impact on user performance. Systematic research, generally of a rather dull sort, can furnish much of the needed information. Such hard information is needed to support a wide variety of categories of requirements. The efforts to refine traditional RGL formulas seem disproportionately large, and foster the continued misuse of an index that is of doubtful applicability and that, outside of a small research community, may be almost universally misinterpreted.

READING PROBLEMS WITHIN THE ARMY

Clinton L. Anderson
Major, General Staff
Education Directorate
Headquarters, Department of the Army
The Adjutant General Center

ABSTRACT

The soldier's inability to read adversely affects the Army mission and personal welfare of individual service members. Recent Army research has revealed that there are significant numbers of poor readers both on active duty and entering the force structure. To help resolve the problem, the Army is conducting a three-pronged attack. First, Army publications will be written and edited to match reading capabilities of their intended users more closely. Second, HQDA is assisting the Office of the Secretary of Defense in striving for workable preenlistment basic skills education programs in the civilian sector sponsored by the Department of Labor and Health, Education, and Welfare. An Army goal is not to enlist any person having less than a 5th grade level in reading, speaking, and listening in English, and in basic arithmetic. Third, the Army is implementing an on-duty Basic Skills Education Program designed to remediate and enhance educational skills needed by soldiers to perform their military occupational specialties and grow professionally within the service. All three of these efforts rely heavily on the individual's ability and motivation to learn and develop skills essential for success within the Army system.

Introduction

The soldiers' inability to read adversely affects the Army mission and soldier welfare. This is substantiated by General Accounting Office (GAO) research, which indicates poor readers have:

1. More disciplinary problems.

2. Higher rates of discharge during and after training.

3. Poor job performance.

4. Higher rates of attrition in technical training.

5. Lack of potential for career advancement.

Good management techniques and personnel administration procedures require that the following steps be taken:

1. Identification of poor readers.

2. Development of educational programs to assist identified deficient readers in reaching requisite reading levels.

3. Close monitoring of the educational programs to ensure that they are effective in both raising the reading levels and eliminating the problems cited by GAO and those identified from within the Army.

4. Assisting Army editors at all levels of command in matching the readability of publications as closely as possible with the reading abilities of the intended users.

Scope of the Problem

Recent Army research has revealed that there are significant numbers of poor readers, both on active duty and entering the Army. First, a representative worldwide sample of 1525 infantry soldiers indicates that approximately 35 percent read below the 7th grade level (Table 9) (LaRocque, 1977). In addition there are significant percentages of poor readers in enlisted ranks E-1 through E-7. Second, a study of 9846 Supply AIT students at the Quartermaster School indicates that approximately 44 percent read below the 7.5 grade level (Table 10) (Hampton, 1977). Finally, testing of accessions into training base indicates that approximately 27.2 percent read below 7th grade level (Table 11) (HQ TRADOC (ATAG-ED), 1977).

Soldiers are required to read increasingly more technical Field Manuals, Technical Manuals, and Soldiers' Manuals. Although every effort is being made to write Army literature at approximately the 7th grade level, this is a Herculean task. Also, in light of the complexity of modern military technology, it is virtually impossible to write below the 7th grade level. In addition, it is readily apparent from the above cited research that writing at the 7th grade level will still exclude large portions of our soldier populations.

All indications point to the fact that reading ability of our manpower pool is decreasing. The best evidence to support this contention is a recent comparison of 1960 and 1972 reading score levels of high school seniors (Beaton, 1977). If you parcel the 1972 population into Quintiles and compare this to the scoring of the 1960 population, a significant shift occurs.

	1972	1960
Top 5th	20%	26.5%
2nd 5th	20%	20.1%
3rd 5th	20%	19.0%
4th 5th	20%	18.3%
Bottom 5th	20%	16.0%

This study may well indicate a trend. Subjectively, at least we can expect a continuing downward trend with relation to accessions.

Given the magnitude of the manpower pools needed in the Army, there is little choice but to accept at least a portion of accessions at less than desirable reading levels. It is only through development of educational programs that the Army can hope to have sufficient productive soldiers to accomplish its mission.

Preenlistment Basic Skills Program

During the review of the Fiscal Year 1978 DoD Budget, the House and Senate Appropriations Committees, in joint session, tasked the Secretaries of Labor and Health, Education, and Welfare to devise programs permitting prospective recruits to obtain essential basic skills prior to their entrance into active military duty.

Table 9

E-1--E-9 by Comprehension

Reading Grade Level	N	CUM N	%	CUM %
		E-1--E-9 (Overall)		
12	566	1525	37	100
11	101	959	7	63
10	69	858	4	56
9	111	789	8	52
8	76	678	5	44
7	61	602	4	39
6	62	541	4	35
5	48	479	3	31
4	167	431	11	28
3	244	264	16	17
2	20	20	1	1
		E-1--E-3		
12	50	328	15	100
11	13	278	4	85
10	8	265	3	81
9	25	257	7	78
8	15	232	5	71
7	15	217	4	66
6	21	202	7	62
5	19	181	6	55
4	61	162	18	49
3	95	101	29	31
2	6	6	2	2
		E-4		
12	173	427	41	100
11	37	254	8	59
10	29	217	7	51
9	32	188	7	44
8	25	156	6	37
7	20	131	5	31
6	18	111	4	26
5	17	93	4	22
4	29	76	7	18
3	43	47	10	11
2	4	4	1	1
		E-5		
12	79	182	43	100
11	14	103	8	57
10	7	89	4	49
9	17	82	9	45
8	5	65	3	36
7	7	60	4	33
6	5	53	3	29
5	0	48	0	26
4	10	48	5	26
3	34	38	19	21
2	4	4	2	2

Note. A 1977 Army-wide study conducted by USAIS, Fort Benning, GA, involving 1525 Infantrymen. Gates-MacGinitie test was used to measure reading grade level. Numbers for individual ranks do not add up to numbers for overall sample because of missing data.

Table 9 (Continued)

Reading Grade Level	N	CUM N	%	CUM %
		E-6		
12	100	195	51	100
11	8	95	4	49
10	19	87	10	45
9	4	68	2	35
8	7	64	4	33
7	6	57	3	29
6	4	51	0	26
5	10	50	5	26
4	0	40	0	21
3	38	40	20	21
2	2	2	1	1
		E-7		
12	50	121	41	100
11	13	71	11	59
10	7	58	6	48
9	3	51	2	42
8	8	48	7	40
7	5	40	4	33
6	2	35	2	29
5	3	33	2	27
4	9	30	8	25
3	19	21	15	17
2	2	2	2	2
		E-8		
12	45	77	58	100
11	3	32	4	42
10	4	29	6	38
9	7	25	9	32
8	3	18	4	23
7	4	15	5	19
6	2	11	2	14
5	1	9	2	12
4	3	8	4	10
3	5	5	6	6
2	0	0	0	0
		E-9		
12	68	110	62	100
11	10	42	9	38
10	5	32	4	29
9	5	27	5	25
8	5	22	5	20
7	2	17	1	15
6	2	15	2	14
5	5	13	5	12
4	7	8	6	7
3	1	1	1	1
2	0	0	0	0

Note. A 1977 Army-wide study conducted by USAIS, Fort Benning, GA, involving 1525 Infantrymen. Gates-MacGinitie test was used to measure reading grade level. Numbers for individual ranks do not add up to numbers for overall sample because of missing data.

Table 10

Reading Level, USAQMS Supply AIT Students

Reading Grade[a] Level	N	%
3.0 to 5.0	1631	16.6
5.5 to 7.5	2681	27.2
8.0 and above	5534	56.2
Total	9846	100.0

[a]Reading grade levels were determined using the reading portion of the USAFI Intermediate Achievement Test (Form C).

Table 11

FY77 Accessions into TRADOC Training Centers

Reading Grade[a] Level	No. Tested	%
0.5 to 5.9	33,949	18.8
6.0 to 6.9	13,133	7.3
7.0 and up	133,048	73.9
Total	180,130	100.0

[a]Best available data provided by TRADOC. Metropolitan Achievement Test was generally used in determining RGL.

> The conferees believe that more effective use of these monies would result from programs that emphasize basic educational skills prior to enlistment. Accordingly, prior to Fiscal Year 1979, the Secretary of Health, Education, and Welfare and the Secretary of Labor, in coordination with the Secretary of Defense, are requested to develop a basic skill program using available resources and expertise. This program will be developed to support the educational competencies required by the military services. (Extract from 4 August 1977 Congressional Record--House, page 8742.)

HQDA strongly supports these initiatives and is prepared to assist the Office of the Secretary of Defense in its DoD role regarding implementing programs of this nature.

Army's Basic Skills Education Program (BSEP)

In full compliance with Congressional guidance, the Army High School Completion Program ceased as an on-duty educational activity effective 30 June 78. The on-duty Basic Skills Education Program is being implemented both in the Army training base and by all major commands during 4th quarter, Fiscal Year 1978. BSEP is the Commander's principal on-duty education program for enlisted personnel. BSEP has been developed by Headquarters, Department of the Army, in conjunction with Headquarters, Training and Doctrine Command (TRADOC), and the American Council on Education, with input from elements within other Major Army Commands. Its three operational phases interlock to form a continuum for the soldiers' career growth. BSEP I, conducted by TRADOC during initial training, will provide soldiers with basic literacy instruction in reading and arithmetic through a 5.0 grade level. Its full implementation is dependent upon resource constraints placed on the Army's training base. BSEP II is conducted by permanent duty stations and is designed to raise educational competencies to a 9.0 grade level. BSEP II is "foundation" instruction that reinforces and develops basic educational skills required in common by most soldiers and is relatable to most, if not all, military occupational specialties (MOS) at the .2 skill level. BSEP III provides functional instruction relatable to specific MOSs or career management fields. BSEP III instruction is beyond the scope of the foundation phase and will include development of educational skills needed for advancement beyond grade E-5, MOS skill level .2. Unlike BSEP I and II, BSEP III will not be fully implemented until 1 January 1979. Implementation guidance for BSEP III is scheduled for distribution to major Army Commands during October 1978. In the interim, Army commanders are authorized to continue existing MOS-related skill development instruction currently provided through the Army Continuing Education System.

Salient features of BSEP made it very attractive for today's Army. First, it tackles squarely soldiers' basic literacy problems (i.e., reading, writing, speaking, listening, and computational deficiencies). Second, it is designed as a command program aimed at helping soldiers perform military jobs more effectively, hence tangible support for unit readiness. Third, it is an Army-wide program with standardized diagnostic testing and entry criteria based on soldiers' actual educational needs. Soldiers' educational credentials, or lack thereof, are not determining factors for participation in BSEP. Fourth, it is conducted during normal duty hours at no expense to the soldier.

BSEP is being tied closely into current Army efforts to correlate reading abilities of soldiers with reading levels of Army publications. The objective of this overall effort is to obtain a high probability that soldiers will be able to read and understand Army publications including training and equipment literature prepared for their use. Army efforts, however, rely heavily on the individual's ability and motivation to learn and develop skills essential for success within the Army system.

Conclusions

In large measure, the Army's problem with reading among its soldiers reflects the larger societal education problem. As weapons systems and tactical warfare have become ever-increasingly sophisticated, the educational capabilities of our soldiers have not kept pace. In fact, a regression may have occurred. Current initiatives are underway to provide a full array of remedial programs as needed to develop and enhance educational skills required to perform military jobs and to grow professionally within the Army. The Army publication system is being fine-tuned to ensure, as closely as possible, that the readability of manuals, regulations, and other written documents is within range of the reading capabilities of the soldiers for whom the material is intended to be used. Soldiers must have positive motivation and good aptitude for learning, however, if these initiatives are to succeed.

DEPARTMENT OF THE ARMY EFFORTS TO IMPROVE READABILITY OF THE ADMINISTRATIVE PUBLICATIONS

Dorothy B. Nicewarner
Department of the Army
The Adjutant General Center, Publications Directorate

ABSTRACT

The Army has launched a program to ensure that administrative publications are written at a reading level appropriate to their users. In January 1978, an editing staff was established and assigned the mission of reducing the reading comprehension levels of Army administrative publications from the 17+ grade level to between the 10th and 12th grades.

The major problem being experienced by the editors is determining the reading grade levels of material accurately. Two readability tests are currently being used in the Army--the Fog Index and the Flesch-Kincaid method. When the same material is tested by the two methods, there is always at least a two-grade variance in the results obtained. Such a variance causes the user to question which method is more accurate.

A study is presently being undertaken to determine whether the Fog Index, the Flesch-Kincaid Readability Formula, or some other as yet unidentified method would be the most accurate and reliable for use Army-wide.

Another point of concern within the editing staff is that comprehension is based on more than sentence length and long words. Although it is felt the logical organization of material is also critical, apparently no method has been devised for measuring this factor. Also, it seems fairly obvious that the graphic presentation of material--including the use of illustrations and tables--contributes to the readability of the material. None of the present readability indexes takes this into consideration. These are two areas that deserve further investigation.

- -

In the newly established editing office, a requirement exists for development of performance standards for the editors. Considering the expertise represented in this workshop, it is hoped some suggestions can be provided on this subject.

The Army has launched a program to tailor Army administrative publications to the needs and skills of those who must read them. In January 1978, the Editorial Control Division was organized in the Publications Directorate within the Adjutant General Center. The mission of the new division is to make Army administrative publications easier to read, to understand, and to use. As part of this mission, the office is also charged with reducing both the number of pages in publications and the number of administrative publications. The following objectives have been established for the editorial staff.

1. Reduce the reading comprehension level of most administrative publications from the current 17+ grade level to the 10th to 12th grade reading level.

2. Reduce the number of pages in administrative publications by 10 percent.

3. Reduce the number of administrative publications by consolidating similar publications when possible.

Because of the nature of material included in Army administrative publications and their primary target audiences, it was determined that the acceptable reading grade level for these publications is the 10th to 12th grade--comparable to the reading level of the Reader's Digest. The 10th to 12th grade level sounds quite high when compared to the 7th to 9th grade limits placed on material contained in training and technical publications. Attaining that goal, however, presents a real challenge to us. Many ARs have been tested for readability and almost all of them are rated at the 17+ grade level. It is realized that using a figure above 17 doesn't really mean anything. Since some Army regulations have been tested out at the 26th grade level, however, reducing that type material to 10th to 12th grade level is a great accomplishment.

At the present time, a contractor is developing a writing and graphic design improvement program for the Army. At the conclusion of the contract, we expect to have, among other things, a writing manual. This manual, which will be distributed to writers Army-wide, will provide a comprehensive document on how to improve one's writing. Another by-product of this contract will be a training package on writing that will be used throughout the Army. Work on the contract is scheduled for completion in January 1979. Mr. Robert Gunning--developer of the Fog Index--is serving as a consultant to the contractor.

The Fog Index is being used to measure the readability of the administrative publications being edited by the Editorial Control Division. The Fog Index was chosen because the formula for its use is so simple--counting words and polysyllables only. In comparing it with other techniques, however (Kincaid, Forcast), we find that the Fog Index always shows the material as being several grades higher than the others. This raises a question as to the accuracy and reliability of the different methods. The one major shortcoming of the Fog Index is that it makes no allowance for the technical, subject-related terms. These are often polysyllables words--but they are well understood. To offset this problem, the Editorial Control Division editors are using their own judgment and omitting from the polysyllable count those words which fit in this category. It has been found that the elimination of the same word a number of times from a passage can result in the reading grade level being lowered by as much as several grades. We don't understand why this is, but it does happen.

I believe an instrument such as the Fog Index should be used as a warning only. I think there is more to comprehension than just sentence length and polysyllables. The organization of a piece of writing is critical--if the material isn't presented in a logical fashion, it can be full of short sentences and words and still make no sense at all. Lack of organization is one of the biggest problems facing our editors. Before they can even attempt to reduce sentence length and eliminate long words, they must put the material in logical sequence. To my knowledge, however, there is no method available for measuring logical organization of material.

I frankly believe that the graphic presentation of a publication--including the numbers and ways in which figures and illustrations are used--makes a difference in the readability of the material. My editors always try to reduce long, complex, narrative material to simple, straight-forward tables and they encourage writers to include illustrations when appropriate. To my knowledge, none of the present readability indexes considers the use of graphics as a factor in determining reading grade level.

At HQDA we have recently drafted a DA Circular on "Reading Grade Levels." In this circular, we have established certain grade levels and testing instruments for the various types of publications. For the present, it appears the Army will be using two different methods for testing--the Kincaid Readability Formula for equipment and training materials and the Fog Index for administrative publications. Our goal is one method that would be appropriate for use Army-wide. It is hoped that this workshop will provide some ideas that will be helpful in determining what the method should be.

As I mentioned earlier, the new Editorial Control Division was established in January 1978. In addition to experiencing difficulty with readability indexes, I am also having problems with establishing performance standards for my editors. During our discussions, this week I hope to gain some information that will assist me in developing such standards. I need to know what is an acceptable rate for the number of pages edited in a given time, how can one distinguish between the different types of manuscripts and editing, etc.

I am very excited about the possibilities this workshop presents to all of us. If I can acquire just a small portion of the information I have come in search of, I will consider it to have been a very worthwhile trip.

TECHNICAL GRAPHICS COMPREHENSIBILITY ASSESSMENT

Thomas E. Curran
Navy Personnel Research and Development Center
San Diego, CA 92151

ABSTRACT

There is virtually nothing in the literature in the nature of empirical evidence as to the parameters of technical graphics that make them more or less comprehensible. This research and development project represents an attempt to develop a methodology for examining graphics in order to identify such parameters and at the same time to attempt to anwer some basic questions about the type of graphics commonly found in technical manuals.

Two types of graphics were selected for use as stimulus materials: an exploded view and a cross-sectional view. Versions of each of these were constructed that had 10, 27, 44, and 62 callouts, respectively. In each of these four conditions for each of the two illustrations, the callouts were arranged on the drawing in sequence (clockwise starting from about 12:00) and in random order. All combinations of these stimuli were presented to 243 subjects (sonar technicians, boiler technicians, and gunners mates), who were asked to locate prescribed information. Certain other features of the callouts, such as circling or not circling, were also manipulated.

The major findings were that much shorter search times were required for callouts in sequence than for those in random orders, and that the times required for searching when callouts were in sequence varied only slightly as the number of callouts increased. The time required when an illustration had 62 callouts was not significantly different from the time required when only 10 callouts were present.

Introduction

It has been reported over and over again that Navy technicians have difficulty reading the materials that they must use on their jobs. There is every reason to believe that, in addition to having difficulties with text with which he must deal, the technician also finds it difficult to use illustrations necessary to his work. Although the problems in this area are mentioned (usually in a role subordinate to that of text) in many, many documents, there has been virtually no empirical evidence presented with regard to the parameters of illustrations or graphics that make them more or less easy to use.[1] The word "readability" does not readily apply to graphics; the word "comprehensibility" does, but a better term is probably "usability." A huge number of sources in computer search libraries list the terms "graphics" or "illustrations" in their titles and/or key words, but to date not one has been found that empirically demonstrates factors affecting their usability.

---

[1]The term "graphics" is taken here to encompass the entire range of pictorials that might be found in technical information. It is intended to include photographs of all kinds, line drawings and diagrams, charts, maps, and tables. Only those portions of technical information that are exclusively text are excluded.

NPRDC has been involved for the past 2 years in attempts to quantify certain of the parameters of graphics that promote their usability and in categorizing the types of graphics so as to enhance communication among users and illustrators. The first thrust, soon to be reported in a NPRDC technical report see Curran & Mecherikoff, 1979 is aimed at the empirical resolution of one small aspect of the overall problem of usability and at the development of a general methodology by which future progress in this area can be made. The second endeavor, which is currently in progress, involves the development of a "taxonomy" of technical graphics that has meaning for all the various persons to whom the comprehensibility or usability of a graphic is important.

Phase One: The Initial Experiment

Rationale

The goal in this phase was to "begin development of empirically-based guidelines and objective measurement techniques to increase the usability of illustrations in technical manuals, thereby reducing the arbitrariness of existing requirements and guidelines" (Curran & Mecherikoff, 1978). It was extremely important that the selection of subject tasks be as realistic as possible, and yet it was considered premature to examine subject performance in an actual job environment in which the technician was using illustrations to do actual maintenance on real equipment. It was also considered vital that the manipulations of the stimulus materials (i.e., the illustrations with which the technician-subject worked) were similar to the type and quality of those he might encounter in the real world. As for the first problem, one of the most common technician activities with regard to illustrations is the location and identification of information. That is, it is not unusual for a person to have a "verbal label" (a number, name, reference designation, etc.) for an equipment part and be required to locate that part physically. Likewise, he must occasionally need to determine the name of a part that he has already located in the equipment and compared with a drawing of that part. These two different tasks point out a major premise of this experiment--that the characteristics of the illustration should be intimately related to the use to which that drawing is being put.

With regard to the selection of the particular illustrations to be used as experimental stimuli, a survey of illustrations existing in operational manuals resulted in a number of candidates for such stimuli (some of which were so poorly constructed as to rule them out even for experimental purposes). From among these candidates, one cross-sectional view and one exploded view were selected as representing typical graphic types and as being amenable to modification to present different "levels" of the variables that the researchers desired to manipulate. This first phase can therefore be summed up as the search for representative exemplars of commonly used technical graphics, manipulation of variables within the selected illustrations, measurement of technician performance that would permit the establishment of certain empirical relationships among drawings, their intended uses, and their critical parameters. This became the prototype for a general methodology by which further studies of a similar type might be conducted.

The Subjects

It was considered that "learning to use illustrations" was an important subject for study, but not one that was of primary consideration here. Therefore, subjects selected for use were Navy technicians in three different ratings, Sonar Technician (ST), Gunners Mate (GM), and Boiler Technician (BT), already having experience in the use of graphics. The ST group represented a population of generally higher ability and of a different orientation (i.e., electronics rather than mechanical) than the latter two groups. In terms of experience, the average times the men had been in the Navy were 3.56, 4.57, and 2.21 years for the ST, GM, and BT groups respectively.

The Method

A technician using an illustration on the job is engaged in an extremely complex chain of behaviors. Examination of the physical equipment, manipulation of tools and test equipment, following of procedural text, and reference to illustrations are but the most visible of these. For the purposes of this experiment, "reference to illustrations" was extracted from this sequence and considered as an integral whole.

The major variables manipulated in the experiment were the number and arrangement of "callouts" on the drawings.[2] The number of callouts ranged from 10 to 62 on each of the 11" by 17" drawings, and the callouts were arranged in either sequential (from roughly 12:00 clockwise) or random order. On the cross-sectional view, one condition involved the use of both numbers and nomenclature (part names) directly on the drawing. Other, secondary conditions involved circling callouts (versus not circling them) and extending the callout leader lines so that the numbers themselves were in straight horizontal or vertical lines. For conditions in which they were required, tables listing the nomenclature of parts in the order of their callout numbers accompanied the drawing itself. Figure 14 shows the cross-sectional view used in the experiment, with both numbers and nomenclature in the callouts (reduced from 11 x 17 to 8½ x 11 page size). Figure 15 illustrates the exploded view (also reduced) used (along with its table), with number callouts in random order, and with leaders extended.

The two characteristics of number and arrangement of callouts were selected as major variables for this study for two reasons. First, more often than not, callouts on illustrations from technical manuals are in semi-random order, which intuitively seemed counter-productive for searching for a specific number.[3] Secondly, the survey of existing technical infomation indicated that not even common sense rules were being followed with regard to the number of callouts on illustrations. Finding graphics similar to that in Figure 16 (originally in 11 x 17 size) was not at all uncommon.

The major performance ("dependent") variable in the experiment was time; i.e., the time required by the subject to point to the part after being given its number or nomenclature, or the time required to verbalize the name of the part when presented with a drawing where the part was highlighted (marked with a red pen). "Correct" or "incorrect" performance, while obviously very important, was not considered a variable in this instance because the subject was told to continue looking until correct performance was achieved.

---

[2] A "callout" is any label or information about a part that appears on a drawing. Callouts may consist of nomenclature, reference designators, numbers keyed to text or tables, or a combination of these.

[3] Military specifications almost always allow for parts on a drawing to be numbered according to "disassembly order." The DoD-wide specification on general requirements for preparation of manuals, for example, states: "Item numbers on exploded views used to show assembly/disassembly shall be in disassembly order." This is in spite of an earlier statement in the same specification which says: "Index (callout) numbers for each separate figure shall start with Arabic number 1 and continue consecutively. Sequence shall be from top to bottom or clockwise, when possible."

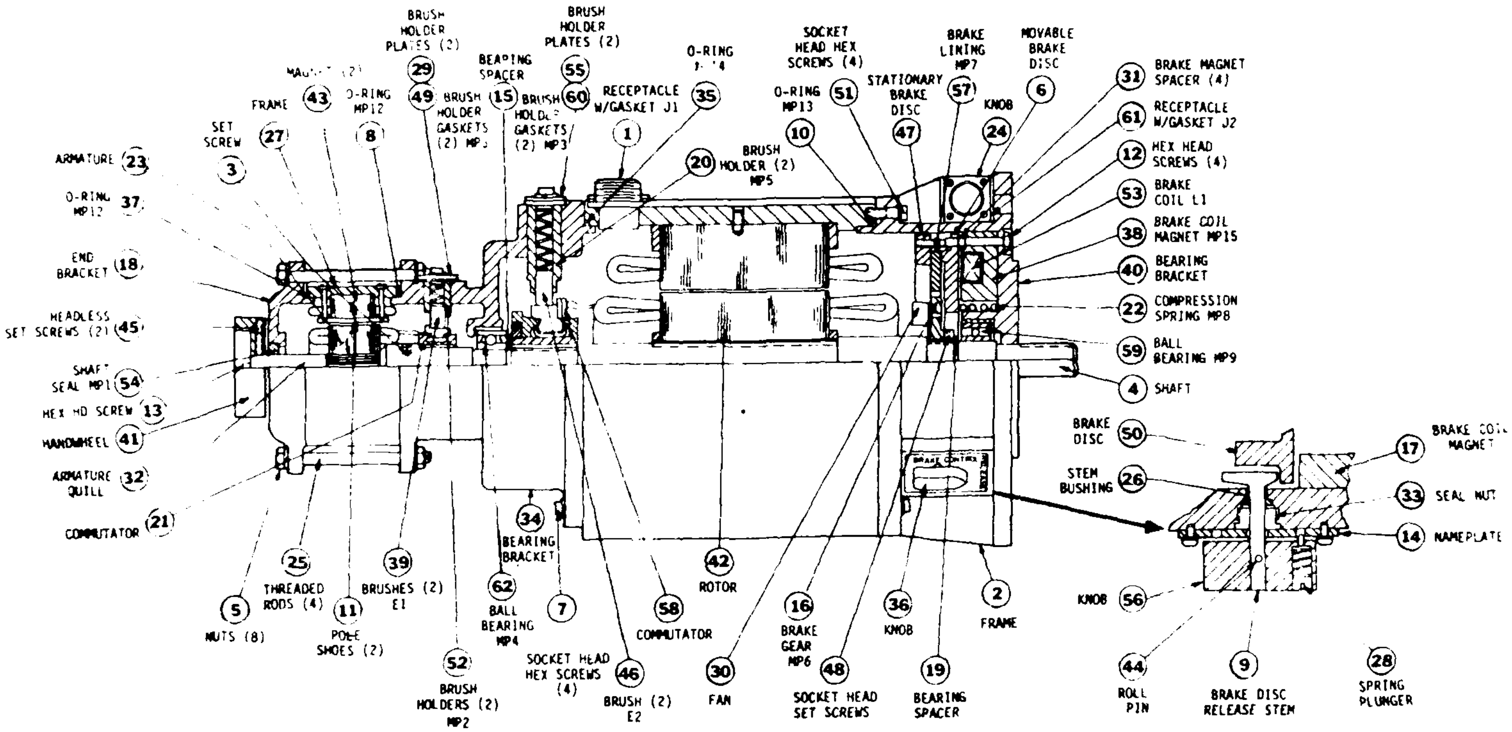
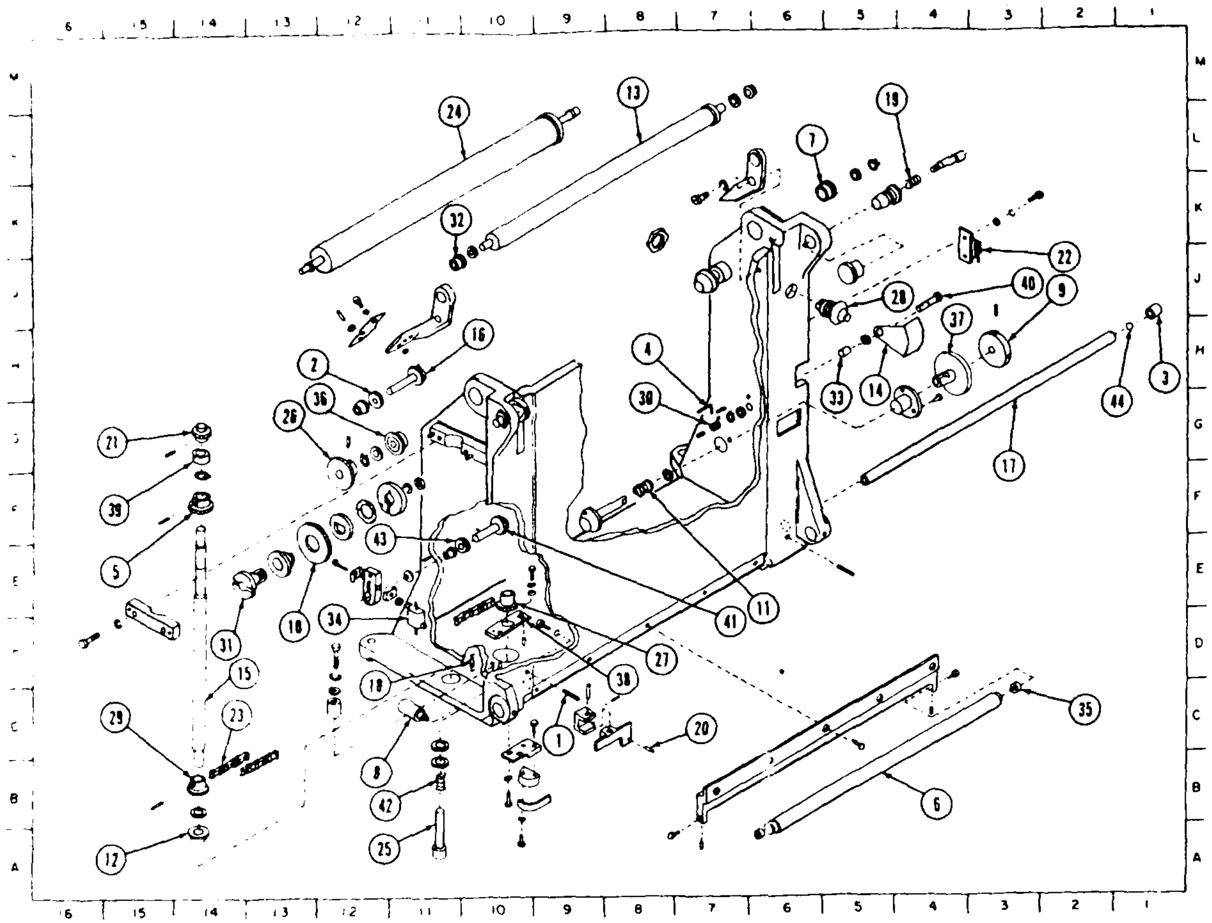


Figure 14. Cross-sectional view used as stimulus in experiment (62 callouts, numbers, and nomenclature, with numbers in random order).



Parts Location Index

CALLOUT NUMBER	REF DESIG	NAME
1	MP1	Spring
2	MP31	Thrust Ball Bearing
3	MP6	Ball Adapter
4	MP11	Paper Measuring Device
5	MP25	Gear
6	MP4	Printing Drum
7	MP20	Flanged Ball Bearing
8	L1	Electrical Solenoid
9	MP8	Knob
11	MP30	Hub Assembly Gear
11	MP10	Compression Spring
12	MP28	Flanged Ball Bearing
13	MP17	Leveling Roller
14	MP13	Dial
15	MP29	Positive Shaft
16	MP32	Hub Assembly Shaft
27	MP7	Tension Roller
18	R1	Resistor
19	MP16	Stud Spring
20	MP2	Scriber
21	MP23	Timing Gear
22	J2	Switch Assembly
23	MP22	Block Chain
24	MP21	Output Roller
25	MP36	Coupling
26	MP19	Roller Shaft Gear
27	MP34	Sprocket Wheel
28	S2	Pushbutton Switch
29	MP26	Sprocket Wheel
30	MP12	Tension Spring
31	MP27	Position Shaft
32	MP18	Flanged Ball Bearing
33	MP15	Bearing
34	CR1	Semiconductor Device
35	MP3	Annular Bearing
36	MP39	Flanged Ball Bearing
37	MP9	Instruction Knob
38	MP37	Adjustment Bearing
39	MP24	Annular Bearing
40	MP14	Shaft
41	MP33	Cap Assembly
42	MP35	Spring
43	MP38	Thrustball Bearing
44	MP6	Ball

Figure 15. Exploded view used as stimulus in experiment, with table (44 callouts, random order, numbers circled, leaders extended)

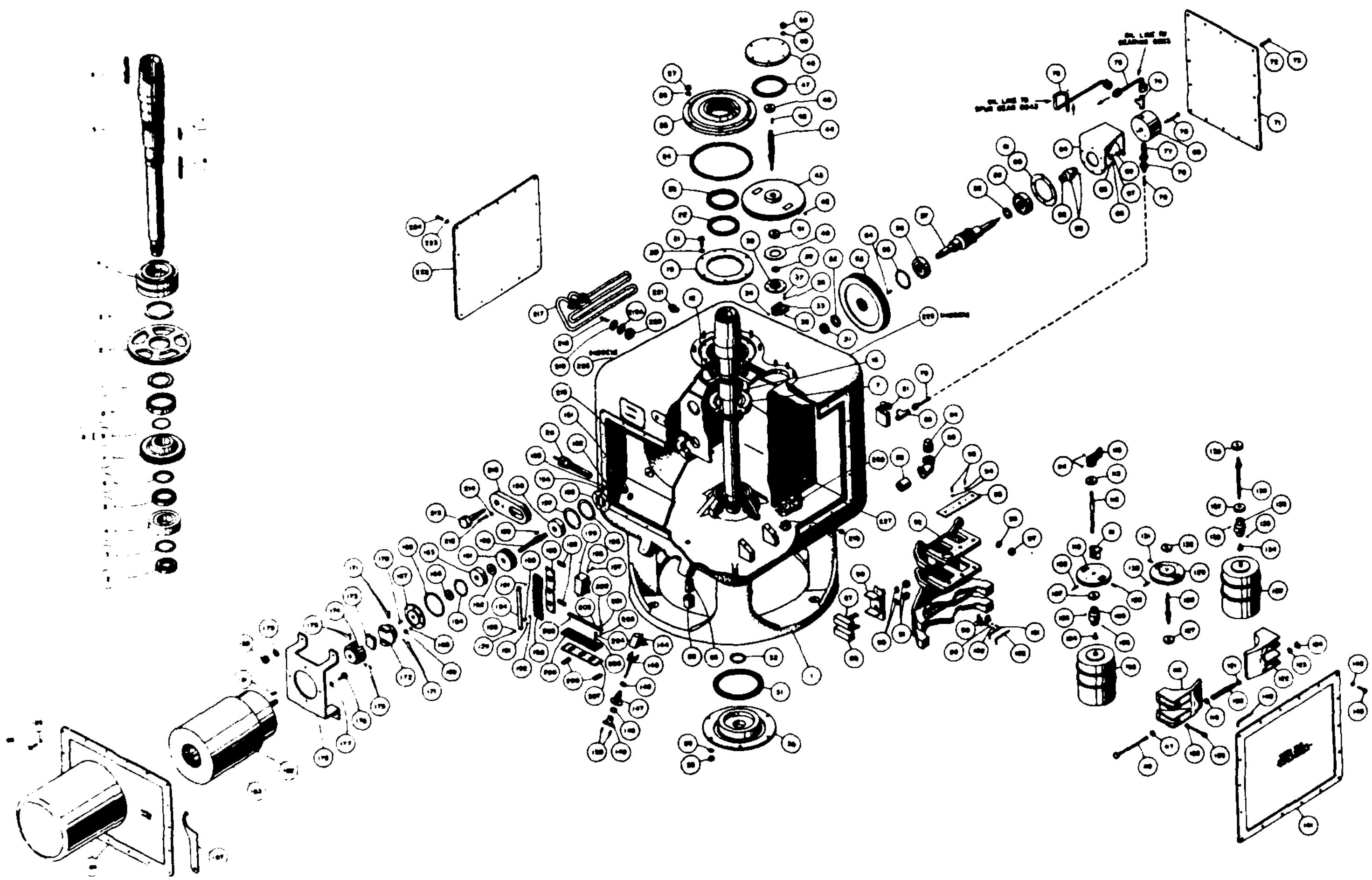


Figure 16. Example of illustration from current technical manual (AN/SPS-10).

The total number of stimulus variations was 160: all combinations of: (1) 10, 27, 44, or 62 callouts, (2) sequence or random order, (3) numbers circled or not circled, (exploded view only), (4) leader lines extended or not extended (exploded view only) and (5) numbers only as callouts versus nomenclature only versus both numbers and nomenclature (on the cross-sectional view only). It was considered too tedious to test each subject on all 160 combinations; therefore, four different groups were presented with 40 different test items each. Test items within each group were representative of the set of all possible combinations of the variables (e.g., each subject saw roughly an equal number of "sequence" and "random" orders).

Results and Discussion

For the purposes of this paper, the results will be presented largely in qualitative form. For the detailed statistics and technical discussion, the reader is invited to examine the forthcoming technical report describing the experiment Curran & Mecherikoff, 1979 .

In general, the major finding with regard to callouts involves the variable of sequence versus random order. It was found that callouts that were numbered in clockwise sequence (counterclockwise would possibly be just as efficient) result in far lower search times than callouts that were distributed in random order. The average times involved in this somewhat artificial situation were relatively short (e.g., 6 seconds as opposed to 2 seconds, let us say). When one considers that the activity of searching for a part, given a callout number, in order to match the picture of that part with the physical object, may occur literally dozens of times in a typical assembly task, however, the cumulative time becomes significant. Further, the subjects exhibited, and often expressed verbally, a much more relaxed and anxiety-free "feeling" when the callouts were in sequence. There is evidence that frustration with poorly organized and constructed technical materials can lead to complete abandonment of the manual. Freedom from such frustration may be a very important element in the degree to which manuals are used effectively.

Highly significant differences were found in many tasks based on the number of callouts involved, but these differences interacted strongly with the sequence-random variable. In brief, when the callouts were in random order, there was a predictably steady increase in search time as the number of callouts increased. When the callouts were in sequence, however, the increase in search time even from 10 to 62 callouts generally was not significant. This, of course, strongly supports the conclusion that callouts should always be in sequence, and that if they are in sequence, the upper limit on number of callouts certainly exceeds 62--an important factor in considering cost-effectiveness.

A number of differences (for the most part so far unexplained) were found between the ST group and the GMs and BTs. These differences were found in what might be considered as "routine" search tasks where the target information was readily accessible and also in certain other tasks in which features of the stimulus materials made the task more difficult for all subjects than would have been expected a priori. In the latter situation, it was often the case that the GM/BT groups exceeded the time required by the STs by large amounts even when search time for the latter group was larger than would be expected. An example of this is present in Figure 14 in which the instructions to the subjects were to "find and point to the part called 'the fan'." Once one has located the fan on the drawing, it tends to stand out from its surrounding, but subjects in general, and the GMs and BTs in particular, took a much longer time to locate it initially than one would predict.

In instances when the task also required location of a part given nomenclature, the GM/BT groups tended to take a longer time, indicating that there may have been an effect based on reading ability. Individual reading test scores were not available, so this possibility could not be more thoroughly tested.

Conclusions

The overall conclusions of the study described above are as follows:

1. When the technician task calls for parts to be located given callout numbers, the callouts should be arranged in sequence beginning at a convenient location near the top of the drawing. The results indicate that this is the optimal procedure even when the technician must follow a procedural text in which the parts are referred to by callout number for assembly or disassembly.

2. When the technician must use the drawing to locate a part knowing the nomenclature of that part, it is efficient to include the nomenclature itself in the callout, at least when the number of callouts is relatively few. While not directly tested in this experiment, it is considered that, when the number of callouts exceeds about 20, the optimal procedure is to put the nomenclature in an accompanying table in alphabetical order and cross-reference to callout numbers that are in sequence on the drawing itself.

3. In either of the above two situations, *if* the numbers on the drawing are in sequence, the uppermost limit on the number of callouts that can be included on the drawing (within the limits of legibility) is unknown, but must certainly be in excess of 62.

4. More research is needed to determine more specifically the unique requirements of different user populations, such as the STs and the GM/BT group used as subjects here. Once differences between such groups have been delineated, specific guidelines responsive to their particular needs must be proposed.

Phase 2: Taxonomy of Graphic Types

Rationale

This phase of the work is now in progress. A major problem in any discussion of graphics is the difficulty in finding a "common ground" among the three principles: the program manager, the illustrator, and the user. George A. Magnan, a foremost expert in the field of technical illustrating, makes this point when he says:

> For any type of visual communication to be really effective, there must be understanding between the three parties involved--the initiator (engineer, executive, or other person who generates the need to express his ideas visually), the illustrator (art director, draftsman, designer, illustrator, and others who translate the initiator's message into pictorial form), and the user (all those who need to understand the pictorial message in order to carry out their own work). (Magnan, 1974, p. 8)

To improve the lines of communication referred to by Magnan, an attempt is being made to organize the multitude of types of graphics into a coherent whole. The initial step in this endeavor was to survey the literature of graphic arts textbooks, military specifications and standards, and military and civilian style guides to whatever extent they were available in libraries. This proved to be an onerous and frustrating task. Virtually no two of the scores of sources examined used the same schema for categorizing types of graphics and it was often difficult to discern when different terms actually referred to the same general type of graphic. The project was then organized into several strands that might yield some consensus and that could run concurrently. Each of these is discussed in some detail below.

Existing Schemata or Taxonomies

This title is somewhat misleading. As nearly as the author can determine, there is presently no taxonomy of graphic types (in the true sense) in existence. We say this, because, by its very nature, a taxonomy must have an algorithm, or key, according to which each instance of a graphic can be assigned to one, and only one, category in the structure. Except when referring to our ultimate goal, therefore, we will use the word "schema" or "structure" instead of "taxonomy," implying that various organizations exist but that these are without the rules by which instances of graphics can be reliably assigned to one or another category.

The basic schema for this work was one that was developed at the outset of our examination of technical graphics. This will be referred to as the NPRDC model, and will be a starting point and basis of comparison with other schemata, but will almost certainly (because of our naivete at the time of its development) bear little resemblance to the taxonomy finally achieved. An important facet of the NPRDC model, and one which remains a central concept, is that illustrations vary on the dimension of "distance from reality," and that this dimension is useful in the consideration of graphic comprehensibility. The foundation for this concept was simply that the closer a graphic represented reality (with the closest being the photograph), the less concern there would be for manipulation of features that enhance its comprehensibility. Now, while this precept is generally true, the intended use of the graphic must again be kept in mind. For familiarizing an untrained technician with a piece of equipment (in the absence of the equipment itself), the photograph may well be the most comprehensible medium possible; for another technician who has the task of disassembling that equipment, the photograph may be much less usable than, for example, an exploded view.

The schema toward which we are working--and ultimately the taxonomy itself--must take into account all the attributes discussed or alluded to above. It must be comprehensive; each instance of a graphic found in technical information must have its place in the structure. Its cells, or categories, must be mutually exclusive; a given illustration should ideally fit into one and only one. (This is an idealistic goal, in that few, if any, taxonomies in nature fulfill its requirements.) And related to the foregoing requirements, the categories themselves must be selected on the basis that all instances of a graphic type are subject to the same type of usability enhancement as all others. That is to say, one would presumably "do" different things to improve the usability of an electronic schematic diagram than to improve a 3-dimensional plane view of an object. Ideally, then, a category would contain as many different types of illustrations as possible (for the sake of efficiency), all of which could be improved in usability by manipulations of essentially the same kind.

Using the NPRDC classification schema as the starting point, then, new notions were incorporated and old ones revised or discarded as other schemata were discovered in the literature. Of the many new schemata that were found, among the most promising were those contained in Military Standard 100B (given in Figure 17) and in the textbook "Technical Illustration" by George Magnan. An amalgamation of these schemata and others provides the vehicle for the second strand in the project (as discussed below) while further refinement of the hybrid model continues.

Survey of Current and Widely Used Technical Manuals

This step was very difficult to begin, but once begun became merely tedious. It involved the use of a checklist composed of the amalgamation of graphic schemata discussed above and the cataloguing of the extent to which each of those types of graphics appeared in manuals currently in use. The manuals to be so catalogued were selected on the basis of a plan for potential use of the data resulting from this step. Once a survey of the types of graphics currently in use has been made, a logical follow-on step would be a determination of the degree to which such illustrations are found to be appropriate and useful by the technicians. The types of equipments carried aboard ships in the San Diego area were catalogued to ensure that, should this step be undertaken, the manuals surveyed would be in relatively wide use and in use by personnel to which the researchers had access. The result was a representative list of equipment types, across electronics, weapons, engineering, and deck categories, which (1) make use of technical information on a relatively wide scale, (2) were common to a relatively large number of different types of Fleet units, and (3) were accessible to the researchers involved. Little progress can be reported on the results of this task at the moment except to say that it is proceeding steadily.

Survey of Industry

Private companies providing equipment for the Armed Forces have, in many cases, fairly specific directives relating to the graphics in the manuals that were wedded to their hardware. Therefore, a mail survey was instituted in which approximately 50 persons representing about 40 different parent companies were queried with regard to the concept of a taxonomy in general and to their individual company guidelines in particular (as they deal with the categorization of graphics). Response to the survey letter has so far been excellent, with little reluctance to point out flaws in our preliminary model, or to the difficulties one can expect to encounter in bringing together the many disciplines and idiosyncratic techniques involved. It is anticipated that the incorporation of comments and suggestions received as a result of the survey into the preliminary model and literature search having gone before will lead to a satisfactory and generally acceptable end product.

Summary

The strands of the current phase of the project--searching the literature for inputs to a taxonomy, cataloguing the types of graphics currently in use, and surveying private industry for suggestions as to the refinement of the preliminary model--will eventually lead to a true taxonomy with an appropriate key or algorithm for assigning instances of graphics to the model. Iterations of the procedure may well be required to accomplish this goal. The cataloguing of graphics in current use, for example, may well have to be repeated when a more acceptable model is available. In the short term, it is clearly hoped however, that even an interim schema will promote better understanding between the user, the illustrator, and the program manager or engineer.

GRAPHICS CLASSIFICATION SCHEMA

CATEGORY OF GRAPHIC

I. PHOTOGRAPH

II. LINE DRAWING

A. DETAIL DRAWING: Depicts complete end-item requirements for the part(s) on the drawing.

1. Monodetail
2. Multidetail
3. Tabulated Detail

B. ASSEMBLY DRAWING: Depicts the assembled relationship of two or more parts.

1. Detail Assembly
2. Installation Assembly
3. Exploded Assembly

C. CONTROL DRAWING: Discloses configuration and configuration limitations, performance, weight, space, etc.

1. Interface Control
2. Installation Control

D. INSTALLATION DRAWING: Shows general configuration and complete information necessary to install item.

E. ELEVATION DRAWING: Depicts vertical projections of buildings or structures or profiles of equipment.

F. CONSTRUCTION DRAWING: Delineates the design of buildings, structures, or related construction.

III. DIAGRAMMATIC DRAWING: Delineates features and relationships of items forming an assembly or system by means of symbols and lines.

A. SCHEMATIC
B. CONNECTION OR WIRING DIAGRAM
C. INTERCONNECTION DIAGRAM
D. LOGIC
E. PIPING

Figure 17. An example of a graphics classification schemata (From MIL-STD-100B).

RECAPITULATION

George R. Klare
Department of Psychology, Ohio University
Athens, OH

As we have seen, the major themes to emerge from the workshop were not really all that new. I list seven below, borrowing from Tom Curran's summary list, and almost every one has been lurking somewhere in the readability literature.

1. Reading Grade Level (RGL)--What does it mean?

2. Text editing--Why not a tri-service system?

3. Technical Reading Ability--Does it modify comprehension?

4. Performance Criteria--Why aren't they used more?

5. Readability Formulas--How might they be improved?

6. Writers Guides and Manuals--They are needed, but are they used?

7. Comprehension--Can't we improve this wobbly keystone of the arch between writer and reader?

That is not to say that the emphases or slants were old, or that they had been solved prior to the workshop. You apparently can't have generals and admirals pushing something like readability with imperatives like "All writing shall be . . .," or "All writers shall write at . . .," or "All readers shall comprehend. . ." without creating some new stress in the system. Perhaps that is not <u>all</u> bad, at least if some needed qualifications can be introduced into the system as controls. Perhaps, that is, if the new emphasis makes writers think more about the difficulty of their writing without at the same time inducing them to "write to formula."

Just as the themes are not new, they are not independent of each other. Writing about one theme necessarily involves comments about one or more of the others. And even more important, writing about one in the Navy may have implications for the Air Force and applications for the Army, or the reverse or obverse. Which, of course, was a major impetus for the workshop.

Because of both characteristics of the themes, familiarity and interdependence, some suggestions seem in order before commenting on the themes themselves.

1. A newsletter is needed. Comments at the workshop seemed to agree on this matter, and the formidable hurdle of finding an editor, etc., should not be allowed to stand in the way. The name "Milestones" was suggested, and "Meterstones" proffered as a more modern alternative (which did not, unfortunately, meet with the degree of agreement anticipated due to its originality).

2. A follow-up conference is needed, and fairly soon. On this matter, again, much agreement came forth. Since the Air Force hosted the first conference, perhaps the Army or Navy might vie for the honor of hosting the second. And, since one section of

the U.S. was represented, perhaps another should be chosen the second time around. What I'm really trying to say is something like "January or February would be good times of the year, and southern California or southern Florida good places."

3. Closer coordination and/or cooperation is needed in research. This is an "apple pie, mother, patriotism" comment--everyone can agree on it. Unfortunately, it is a tough matter to put into effect. I suggest the establishment of a review board consisting of equal numbers of Air Force, Army, and Navy members. I suggest that, when a piece of research is proposed to that board by one service, the representatives of that service be doubled for that occasion. I am not suggesting that the review board ever be empowered to actually vote down a proposal, but almost certainly the comments that emerge will often aid the research. And, I strongly suspect, a research will go against those occasional strongly and/or uniformly negative reactions only with caution. I see several potential values to the above arrangement.

1. Most obvious, some overlap might be avoided.

2. Perhaps not quite so obvious, some research might be broadened enough to provide for the needs of one or both other services.

3. Still less obvious, but not less desirable, is the possibility of a pool of joint funds for some basic research in literacy or readability. If the services are going to continue to be involved in these fields as it now appears, such research is needed. Otherwise, progress beyond present knowledge becomes unlikely. The same issues and data will be offered in slightly modified form, and worse, misuses will become indistinguishable from legitimate uses. This is happening now even for users with scrupulous intent--to say nothing of those with unscrupulous intent.

But, back to the themes themselves. No attributions have been given for the ideas, since (except for those specifically addressed by working papers) that would have been impossible.

1. Reading Grade Level (RGL)--what does it mean? We need a good statement of how formulas assign grade levels, and how they came about in the first place. Does it really mean anything to say a piece of writing is "22nd reading grade level," as some formulas predict? We also need a good statement of the accuracy of an assigned grade level--perhaps several, one statistical and one individual. The former could well be a statement of the standard error of estimate of a given formula's scores; and the latter, of how level of motivation or of prior knowledge, for example, affect the accuracy of a formula's grade level predictions. Two big bugaboos could be tackled. First, does reading grade level mean the same thing under real life conditions (i.e., in the field, with unobtrusive measures) as it does in the laboratory (i.e., in an experiment, with highly obtrusive tests or other measures)? Second, does reading grade level mean the same thing when reading to learn versus to perform a particular activity versus to look up items of information?

2. Text editing--Why not a tri-service system? Here is one area where some uniformity between the services could be very helpful. Considering the number of pages involved (the Navy is said to have 25,000,000 pages of technical information alone), the savings could be considerable within and between the services. And some nice touches could be passed around: computerized readability information, perhaps even on-line

programs for automatic word-frequency counts; use of the author-aids being developed to help in writing readably of the special controlled vocabulary systems (i.e., Caterpillar Basic English or ILSAM) now available. Work is being done in text editing—why not share it and develop the best possible system?

3. Technical Reading Ability--Does it modify comprehension? Yes and No. Both answers seemed to emerge from the workshop discussion, at least. A number of conferees felt certain that the special knowledge of the experienced person would show up as superior ability to read a document in his/her area. And one study showed that removing "prior knowledge" from subjects' multiple-choice comprehension scores led to greatly improved correlations with both Cloze scores and readability measures. This interpretation did not seem to fit what appeared to be a similar study too well, however. And a study in progress suggested that those "terms everybody in the field would know" are not, in fact, known by a very large percentage of such readers. We await the final results of the study, with the expectation that, if the preliminary findings hold up, some will disbelieve them to the extent of running their own studies. Good! We can only hope this happens, and that several kinds of relevant performance criteria are used as dependent variables. Which brings us to the next theme.

4. Performance Criteria—Why aren't they used more? A simple first answer is that they are harder and less convenient to use than verbal comprehension criteria (see theme 7 below). But studies of this sort are, fortunately, getting more common. A very good example was presented for learning from computer-assisted instructional materials. And the study employed another procedure that is used much too infrequently: unobtrusive measures. If use of laboratory techniques can cause problems of interpretation in physics, it seems surprising that this problem is ignored so often in psychology, when studying the behavior of human subjects in a task as complex as "comprehending." The results of readability studies might well be more persuasive if the above two considerations, use of relevant performance criteria and use of an unobtrusive approach, were given more consideration.

5. Readability Formulas--How might they be improved? Several conferees mentioned plans to develop new or to improve existing readability formulas. The proposals differed somewhat, as is to be expected and encouraged. The literature does, however, suggest certain paths unlikely to lead to much success. Perhaps chief among these is to search for better index variables for predicting readability. This is not, of course, meant to apply to restandardization of existing formulas for specific groups (e.g., enlisted men or recruits, skilled technicians, etc.). Even here, however, going too far in the direction of specialized formulas for special groups necessarily means restricted applicability. Also, the direction of looking at word frequencies or familiarities within an area, promising as it can be for writers (i.e., production) may be of limited value for readability measurement (i.e., prediction). A look at essentially fruitless efforts to improve the Dale list for specific audiences should give one pause. But, in the end, it is an empirical matter, and we need more research here.

6. Writers' Guides and Manuals--They are needed, but are they used? There was a good deal of agreement that writing guides and/or manuals are needed in order to help writers in the military services. In some cases, the writers are not really "writers," but instead are technical experts (if that). In such cases, the need is for rather basic writing skills (apparently the Air Force has a training course for writers). But even where the writers are more skilled, there is still a need for information about writing readably. And with the pressure being put on editors by generals and admirals ("All writing leaving this office shall have a readability grade level no greater than 9"), such guides and manuals should be of great help. Otherwise, editors will be forced to make mechanical changes in

the writing to satisfy a formula score, and this mixing of the prediction/production functions has little chance of successfully improving reader comprehension. A big question that remains is: What kind of guide/manual can be most helpful? Many already exist, and, as an unpublished study shows, they disagree with each other. Two recent manuals, based upon use of research with readers and/or psycholinguistic research, take different tacks. One, Guidebook for the Development of Army Training Literature, provides before-after examples; the other, A Manual for Readable Writing, presents tested principles from research with examples of the kinds of changes that improve readability. And to show the backward state of research in this area, we don't even know how successful either one is, let alone which (if either) approach is better. Guides/manuals are needed, yes. But will they be used?

7. Comprehension--Can't we improve this wobbly keystone of the arch between writer and reader? On the matter of comprehension, how best to define it, let alone how to measure it, can only be described as uncertain, unclear, confused, disagreed upon, etc. at this time. Researchers have been going ahead with the methods available, of course, and this will need to continue. But if ever evidence were needed for a program in basic research, this ought to be enough to convince most any skeptic. Some interesting work is going on now under the general headings of schematics, prior knowledge, and the structure of text, but it has not found its way into the literature of applications. That literature is still digesting the short-term/long-term memory work. As a parting shot (from under the water, on the ground, or in the air), follow-up approaches (the newsletter and more conferences such as this one) should include such new emphases in their agenda along with practical problems of application. With a group of conferees as receptive to new ideas as this one was, this "theoretical stuff" can lead to very practical ideas for application.

AMPLIFICATION OF TRI-SERVICE READABILITY WORKSHOP PRESENTATIONS RELATIVE TO COMMUNICATING WITH THE READER POPULATION

William Adams
Staff, Chief of Naval Education and Training
Naval Air Station
Pensacola, FL 32508

Editor's Note: This paper reflects the views of the writer relative to the Tri-Service Readability and Literacy Workshop. The sponsor of the workshop requested the participants, if they so desired, to submit a paper on pertinent discussion topics soon after the conclusion of the workshop. This paper is an amplification of the points arising at the workshop which it was felt should be included in the workshop proceedings.

Introduction

It is a well recognized fact that a disparity exists between the reading grade level (RGL) of recruits in the armed services and the RGL of technical reading materials. In the 1980s, the armed services will experience fewer qualified military availables because of the decline in the birth rate in the 1960s. Of those qualified military availables, many will lack the basic skills to complete the requirements for military service.

The comments of the participants during the workshop indicated that a reader did not need a 14th RGL to read a manual written at the 14th RGL. The reader may take much longer to read the manual but, if the interest is there, the manual will be completed. The RGL of a manual may range from the 6th to 14th RGLs with an average RGL of 11. What does the RGL mean to the intended user, the producer, and the manager of technical reading materials?

It is suggested that a paper on RGLs be prepared by a recognized member of the tri-service readability workshop. The paper should explain, among other pertinent information, the following: (1) the meaning and use of the RGL, (2) the meaning of an average RGL, (3) the difference between RGL and comprehension, and (4) whether a reader with a 7th RGL can comprehend material written at the 14th RGL under various conditions. It is felt that a paper prepared by a person who has done considerable study relative to RGL and comprehension would be invaluable to personnel in production and management without a background in this area. A paper such as this would assist managers in determining for some situations if a RGL problem really exists. It would definitely eliminate the assumption by managers that a reader must have 16 years of education to read technical material at the 16th RGL.

The presentation of Dr. Tom Curran, NAVPERSRANDCEN, was of particular interest concerning CNET Support Report 2-75. Dr. Curran stated that adjustments should be made to the RGLs of the study because recruits were used in the study that did not possess the prerequisites. Recruits of 6 months were used to determine RGLs of Disbursing Clerks (DK) 1st class and chief that did not possess the vocabulary that would have been acquired by a Disbursing Clerk 2nd class.

The presentation of Dr. J. D. Kniffin, Westinghouse Electric Corporation, on the use of automated publishing, indicated several distinct advantages: (1) automated readability calculations would provide RGLs of random selected passages and an average RGL of the text, (2) the possible use of computerized readability editing to flag unfamiliar words and phrases, and (3) word substitution lists would be provided for substituting more readable words.

Dr. James Burkett, AFHRL Lowry AFB, indicated that writers in the Air Force were writing to established RGLs for Air Force Specialty Codes (AFSC) but were not producing the desired products. Dr. Burkett stated that, if there is a way for writers to get around writing to a RGL formula, they will do so. Dr. Burkett urged that some other method should be identified for producing readable material for the intended population.

Major Mike Birdlebough, Headquarters Air Training Command, commented at the end of the workshop that his interest was in preparing training materials for the lower quality of recruits that will be enlisting in the 1980s. The time to begin preparing such materials is now and not wait until the 1980s. Although the Air Force has established RGLs for AFSCs, Major Birdlebough stated he was in favor of discontinuing the use of RGLs for preparing training materials for the 1980s. The training materials would be written to lower RGLs than those established for AFSCs, and an automated publishing system with some of the characteristics mentioned in Dr. Kniffin's presentation would be used.

The presentation of Dr. Robert Fishburne, Calspan Corporation, on "Validation of Naval Readability Indices," has very good implications. The study was done on programmed instruction material in the Navy's Computer Managed Instruction (CMI) System after implementation into the training environment. The study showed that the RGL of the participants and readability of material were very close and the error rate was low. This study would tend to indicate that the programmed instruction material, if properly developed and validated to established criteria, should be suitable to the RGL of the intended reader population. The Instructional Program Development Centers (IPDCs) are developing instructional materials based on established criteria so the readability of the materials should be compatible with the RGLs of the intended reader population. This type of results can only be obtained by using a cross section of the intended reader population during validation.

The preceding paragraphs emphasize some of the salient discussions that were of direct interest to this CNET representative. CNET was contemplating the use of RGL, when made available, in the development of nonresident training. Dr. Tom Duffy, NAVPERSRANDCEN, stated that he would be publishing the RGLs for clusters of Navy ratings in the near future. Since Dr. Burkett, AFHRL, stated that the Air Force did not experience desirable products as a result of writing to established RGLs, CNET will not request nonresident training developed in-house to be written to established RGLs. RGLs may be helpful to contractors, however, as expressed by Dr. Kniffin, in order to provide readable materials for the intended reader population.

The NAVEDTRAPRODEVCEN is presently writing nonresident training materials to the 7th RGL. The writers use a Thorndike word list based on a 7th grade Thorndike-Barnhart Dictionary. Of course, the technical vocabulary of the specific rating remains unchanged. This approach goes along with Major Birdlebough's (ATC) interest in writing training materials at lower levels rather than RGLs established for a specific AFSC in the Air Force.

CNET is presently looking into the requirements for utilizing an Automated Publishing System for producing rate training manuals, nonresident career courses, advancement examinations, and IPDC instructional materials. The use of this type of system with the characteristics stated by Dr. Kniffin should greatly reduce the production time for the above mentioned products. The average RGL and the range of RGLs can automatically be

AD-A083 043 NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER SAN D--ETC F/G 5/9
TRI-SERVICE LITERACY AND READABILITY: WORKSHOP PROCEEDINGS.(U)
MAR 80 T E CURRAN
UNCLASSIFIED NPRDC-SR-80-12 NL

2 of 2
AD A083043

END
DATE FILMED
5-80
DTIC

provided for each product without the need for additional manpower and resources. Again, Major Bridlebough expressed a desire to use an Automated Publishing System for the Air Force.

The tri-service workshop has confirmed that the approach CNET is presently pursuing is the most appropriate at this time. It is of great benefit to managers, producers, and researchers to participate in periodic workshops of this type so as to minimize the reinvention of the wheel. Workshops of this type are the best means of keeping abreast of the latest developments relative to readability and comprehension.

The personnel of the N-51 branch of CNET are willing to cooperate with other branches of the Armed Services and, where feasible, share expenses in this area of research and development.

REFERENCES

Aiken, E. Personal communication, July 1978.

American Institutes for Research in the Behavioral Sciences. Word frequencies in technical instruction. Prepared for Personnel Research Division, Bureau of Naval Personnel, Department of the Navy, Contract NOp-1498, August 1964.

Barnhart, C. L. (Ed.) The Thorndike-Barnhart comprehensive desk dictionary. Garden City, NY: Doubleday & Company, Inc. 1958.

Beaton, A. E., et. al. Changes in the verbal abilities of high school seniors, college entrants, and SAT candidates between 1960 and 1972, June 1977.

Biersner, R. J. Reading grade levels of Navy rate training manuals and non-resident career courses (CNETS Rep. 2-75). Pensacola, FL: Chief of Naval Education and Training Support, May 1975.

Bluejackets' Manual (20th Edition). Annapolis, MD: United States Naval Institute, in press.

Bormuth, J. R. Readability: A new approach. Reading Research Quarterly, 1966, 1, 79-132.

Braby, R., Parrish, W. F., Guitard, C. R., & Aagard, J. A. Computer-aided authoring of programmed instruction for teaching symbol recognition (TAEG Rep. No. 58). Orlando, FL: Training Analysis and Evaluation Group, in press.

Carroll, J. B. Learning from verbal discourse in educational media: A review of the literature (Research Bulletin RB-71-61). Princeton, NJ: Educational Testing Service, October, 1971, p. 51.

Carver, R. P. Improving reading comprehension: Measuring readability (Final Rep., Contract No. N00014-72-C0240). Silver Spring, MD: Office of Naval Research, American Institutes for Research, 14 May 1974.

Caylor, J. S., Sticht, T. G., Fox, L. C., & Ford, J. P. Methodologies for determining reading requirements of military occupational specialties (Tech. Rep. No. 73-5). Monterey, CA: HumRRO Western Division, Human Resources Research Organization, March 1973.

Chomsky, N. Language and mind (Enlarged Edition). New York: Harcourt Brace Jovanevich, 1972, p. 107.

Coleman, E. B. Developing a technology of written instruction: Some determiners of the complexity of prose. In E. Z. Rothkopf and P. E. Johnson (Eds.), Verbal learning research and the technology of written instruction. New York: Teachers College Press, Columbia University, 1971.

Curran, T. E. Survey of technical manual readability and comprehensibility (NPRDC Tech. Rep. 77-37). San Diego: Navy Personnel Research and Development Center, June 1977. (AD-A042 335)

Curran, T. E., & Mecherikoff, M. Quantification of technical manual graphics comprehensibility (NPRDC Tech. Note 78-2). San Diego: Navy Personnel Research and Development Center, January 1978.

Curran, T. E. & Mecherikoff, M. Toward more comprehensible technical manual graphics (NPRDC Tech. Rep. 79-28). San Diego: Navy Personnel Research and Development Center, July 1979. (AD-A074 967)

Dale, E., & Chall, J. S. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 11-20 and 37-54.

Dale, E., & O'Rourke, J. The living word vocabulary. Elgin, IL: Dome, Inc. 1976.

Department of Defense. Military specification, manuals, technical: General requirements for preparation of (MIL-M-38784A). Washington, DC: Department of Defense, January 1977.

Duffy, T. M., & Nugent, W. A. Reading ability and attrition of June enlistees. Unpublished.

Duffy, T. M. Literacy training in the Navy. In Fletcher, J. D., Duffy, T. M., and Curran, T. E., Historical Antecedents and contemporary trends in literacy and readability research in the Navy (NPRDC Tech. Rep. 77-15). San Diego: Navy Personnel Research and Development Center, January 1977. (AD-A035 582)

Entine, E. B., & Klare, G. R. Some inter-relationships of readability, cloze, and multiple-choice comprehension scores on a reading comprehension test. Submitted for publication. Journal of Reading Behavior.

Fass, W., & Schumacher, G. M. The effects of motivation, subject activity, and readability on the retention of prose materials. Journal of Educational Psychology, in press.

Flesch, R. F. A new readability yardstick. Journal of Applied Psychology, 1948, 32, 221-233.

General Accounting Office. A need to address illiteracy problems in the military service (Rep. No. FDCD-77-13). Washington, DC: United States General Accounting Office, 31 March 1977.

General Motors. STAR: General Motors computerized simple test approach for readability, a tool for improved communications. Undated.

Githens, W. H., Shennum, W. A., & Nugent, W. A. Personnel characteristics relevant to Navy technical manual preparation (NPRDC Tech. Rep. 76-26). San Diego: Navy Personnel Research and Development Center, December 1975. (AD-A019 366)

Greeno, G. Process of understanding in problem solving. In Catellan, N. J., Jr., Pisoni, D. B., & Potts G. R. (Eds.), Cognitive theory. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977, 43-83, p. 44.

Guitard, C. R. Computer-aided authoring system user's guide: Volumes I and II (TAEG Rep. No. 59). Orlando, FL: Training Analysis and Evaluation Group, 1978.

Hampton, R. Comprehensibility study. Fort Lee, VA: US Army Quartermaster School, 1977.

Harris, A. J., & Jacobson, M. D. Basic elementary reading vocabularies. New York: MacMillan, 1972.

Harris, A. J., & Jacobson, M. D. Appendix D, The Harris-Jacobson readability formulas. In Harris, A. J., & Sipay, E. R. How to increase reading ability. New York: David McKay Company, Inc., 1975.

HQ TRADOC (ATAG-ED) 1st Endorsement to HQDA (DAPE-MPE-CS) letter, subject: Information on New Accessions Entering Training, dated 20 December 1977.

Holmes, J. A. Speed, comprehension, and power in reading. In Bliesmer, E. P., and Staiger, R. C. (Eds.), Problems, programs, and projects in college-adult reading. Eleventh Yearbook of the National Reading Conference. Milwaukee: The National Reading conference, Inc., 1962, 6-14, pp. 9-10.

Huey, E. B. The psychology and pedogogy of reading. New York: The Macmillan Co., 1908, pp. 147-148.

Hughes Aircraft Company, Ground Systems Group. Task 1 report: Analysis of current and proposed technical manual systems. Prepared for the David W. Taylor Naval Ship Research and Development Center, Bethesda, MD: 1977.

Hughes Aircraft Company. Preliminary NTIP System concept and alternative configurations: Addendum--concept of the user-data match model. For David W. Taylor Naval Ship Research and Development Center, Bethesda, MD: (Contract N00600-76-C-1352), January 1978.

Keeler, F. L. An automated publishing system for the Naval Education and Training Command (TAEG Rep. No. 50). Orlando, FL: Training Analysis and Evaluation Group, 1977.

Kearn, R. P., Sticht, T. G., Welty, D., & Hauke, R. N. Guidebook for the development of Army training literature. Special publication, U.S. Army Research Institute for the Behavioral and Social Sciences. Alexandria, VA: Human Resources Research Organization, November 1975.

Kincaid, J. P., & Fishburne, R. P. Readability formulas for military training materials, Human Factors Society Bulletin, 1977, 20, (7).

Kincaid, J. P., Rogers, R. L., Fishburne, R. P., & Chissom, B. S. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel (Branch Report 8-75). Millington, TN: Chief of Naval Technical Training, Naval Air Station, 1975.

Kincaid, J. P., Yasutake, J. Y., & Geiselhart, R. Use of the automated readability index to assess comprehensibility of Air Force technical orders (SEC-TR-67-47). Wright-Patterson AFB, OH: Aeronautical Systems Division, 1967.

Kintsch, W. The representation of meaning in memory. New York: Wiley, 1974.

Klare, G. R. The measurement of readability. Ames, IA: Iowa State University Press, 1963.

Klare, G. R. A manual for readable writing. Glen Burnie, MD: REM Co., 1975.

Klare, G. R. A second look at the validity of readability formulas. Journal of Reading Behavior, 1976, 8, 129-152.

Klare, G. R., Rowe, P. R., St. John, M. G., & Stolurow, L. M. A first version of a computer-aided revision, editing, and translation system (CARET I) (Tech. Rep. No. b.2.1). Cambridge: Harvard University Computer Aided Instruction Laboratory, 1969.

Kling, M., & Pratt, D. F. Readability: A re-examination and new directions for further research. Paper presented at the annual meeting of the National Reading Conference, New Orleans, 2 December 1977.

Kucera, H., & Francis, W. N. Computational analysis of present-day American English. Providence, RI: Brown University Press, 1967.

La Rocque, R. Comprehensibility study. Fort Benning, GA: U.S. Army Infantry School, 1977.

Locke, J. Of the conduct of the understanding, 1697. In F. W. Garforth (Ed.), John Locke's of the conduct of understanding, New York: Teachers College Press, 1966, p. 73.

Magnan, G. A. Using technical art, an industry guide. New York: Wiley-Interscience, 1970.

New York Times News Service. Press release dated 16 November 1977.

Post, T., Price, H., & Diffley, G. A guide for selecting format and media for presenting maintenance information. BioTechnology Inc. for David W. Taylor Naval Ship Research and Development Center, Bethesda, MD: Contract N00600-76-C-1373, November 1976.

Powers, T. E. Navy enlisted personnel characteristics--preliminary analysis. ManTech Corporation of New Jersey, Rockvill, MD. fpr Navy Technical Information Presentation Program, Contract N00600-76-D-0306, June 1976.

Powers, T. E. Selecting presentation modes according to personnel characteristics and the nature of job tasks, Part I: Job tasks. Baltimore: University of Maryland, for Navy Technical Information Presentation Program, Contract N00014-76-C-1067, January 1977.

Sheffield, A. D. (Ed.) Soule's dictionary of English synonyms. Boston: Little, Brown, and Co., 1959.

Simons, H. D. Reading comprehension: The need for a new perspective. Reading Research Quarterly, 1971, 6, 338-363, p. 340.

Skinner, B. F. Verbal behavior. New York: Appleton-Century-Crofts, 1957, p. 277.

Smith, E. A., & Kincaid, J. P. Derivation and validation of the Automated Readability Index for use with technical materials. Human Factors, 457-464, 1970.

Spache, G. D. What is comprehension? In E. P. Bliesmer and R. C. Staiger (Eds.) Problems, programs, and projects in college-adult reading. Eleventh Yearbook of the National Reading Conference. Milwaukee: The National Reading Conference, Inc., 17-19, 1962, p. 17.

Stocker, L. P. Increasing the precision of the Dale-Chall readability formula. Reading Improvement, 1971-1972, 8, 87-89.

Szalay, T. G. Validation of the Coleman readability formulas. Psychological Reports, 1965, 17, 965-966.

Thorndike, E. L. The understanding of sentences. Elementary School Journal, 1917, 18, 98-114, p. 114.

Thorndike-Barnhart handy dictionary. Garden City, NY: Doubleday and Company, Inc., 1955.

United Press International. Press release dated 25 August 1977.

DISTRIBUTION LIST

Chief of Naval Operations (OP-102) (2), (OP-11), (OP-987H)
Chief of Naval Research (Code 450) (3), (Code 452), (Code 458) (2)
Chief of Information (OI-2252)
Director of Navy Laboratories
Chief of Naval Education and Training (00A), (N-5)
Chief of Naval Technical Training (Code 017)
Commander Training Command, U.S. Pacific Fleet
Commander Training Command, U.S. Atlantic Fleet (Code N3A)
Commander, Naval Military Personnel Command (NMPC-013C)
Commander, David W. Taylor Ship Research and Development Center (1803) (15)
Commanding Officer, Fleet Combat Training Center, Pacific (Code 00E)
Commanding Officer, Naval Education and Training Program Development Center (Technical Library) (2)
Commanding Officer, Naval Development and Training Center (Code 0120)
Director, Training Analysis and Evaluation Group (TAEG)
Master Chief Petty Officer of the Force, U.S. Atlantic Fleet
Master Chief Petty Officer of the Force, U.S. Pacific Fleet
Master Chief Petty Officer of the Force, Naval Material Command (NMAT 00C)
Master Chief Petty Officer of the Force, Naval Education and Training Command (Code 003)
Manpower and Personnel Division, Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base
Technical Library, Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base
CNET Liaison Office, Air Force Human Resources Laboratory, Williams Air Force Base
Technical Training Division, Air Force Human Resources Laboratory, Lowry Air Force Base
Army Research Institute for the Behavioral and Social Sciences (Reference Service)
All Workshop Participants
Defense Technical Information Center (12)