DOCUMENT RESUME

ED 074 472 CS 000 448

AUTHOR Carver, Ronald P.
TITLE New Techniques for Measuring and Improving Reading Comprehension. Technical Report No. 1.
INSTITUTION American Institutes for Research in the Behavioral Sciences, Silver Spring, Md.
SPONS AGENCY Office of Naval Research, Washington, D.C. Personnel and Training Research Programs Office.
REPORT NO TR-1
PUB DATE Feb 73
NOTE 89p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Cloze Procedure; *Computer Assisted Instruction; Content Reading; Objective Tests; Programed Instruction; *Reading Comprehension; *Reading Research; Reading Skills; *Reading Tests; Secondary School Students; *Testing

ABSTRACT

A standardized method has been developed which will convert prose training materials into a form which forces the trainees to read the material with at least a minimal level of comprehension. The materials, called programed prose materials, are developed in an objective manner amenable to computer production. Phase I of the project involved an extensive investigation of a new technique, the reading-storage test, for measuring the learning that occurs during reading. The results suggested that the completely objective reading-storage test provides a better measure of the primary effects of reading than its two closest rivals--the cloze test and the paraphrase test. In the Phase II experiment, programed prose was compared to regular prose under low and high motivation conditions. The programed prose facilitated learning under the low motivation condition and inhibited learning under the high motivation condition. Phases III and IV, to be conducted within the next year, will assess the effectiveness of programed prose of varying levels of material difficulty and will investigate the effectiveness of the method in a situation where it would be expected to be maximal. (Author/TO)

ERIC

ED 074472

New Techniques for Measuring and Improving Reading Comprehension

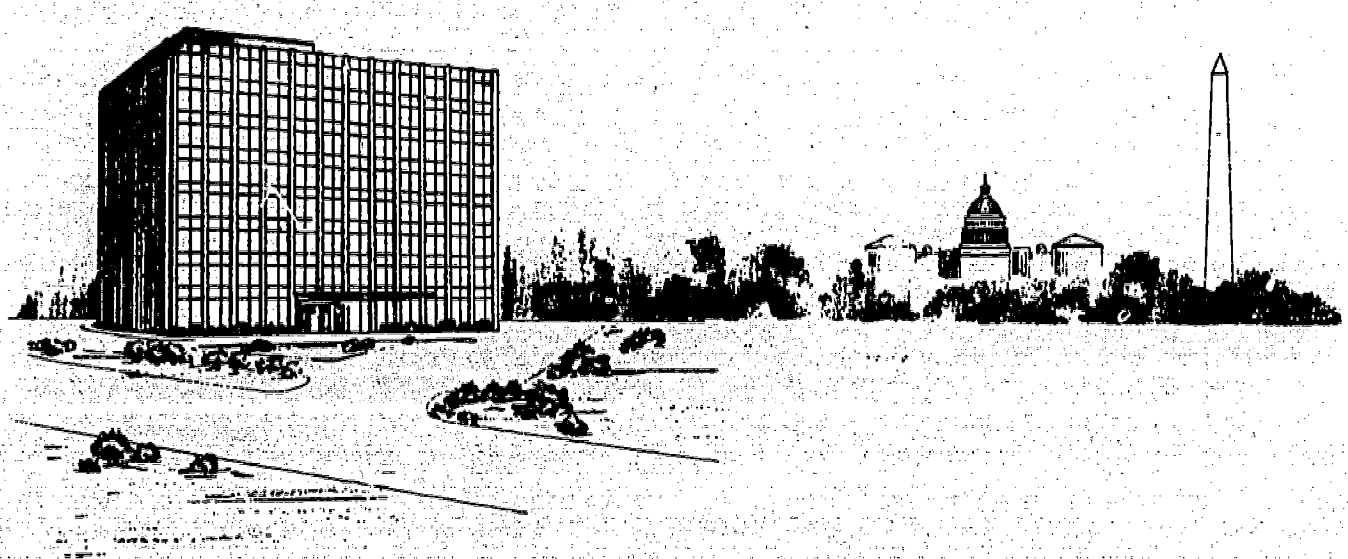
U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

SCOPE OF INTEREST NOTICE
The ERIC Facility has [illegible] this document for [illegible] to [illegible]

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

Ronald P. Carver

Technical Report No. 1
FEBRUARY 1973



CS 000 448

AMERICAN INSTITUTES FOR RESEARCH
WASHINGTON OFFICE
Address: 8555 Sixteenth Street, Silver Spring, Maryland 20910
Telephone: (301) 587-8201

R73-2

AMERICAN INSTITUTES FOR RESEARCH

WASHINGTON, D.C.

EDWIN A. FLEISHMAN, PhD, DIRECTOR

Albert S. Glickman, PhD, Deputy Director

INSTITUTE FOR COMMUNICATION RESEARCH

Arthur L. Korotkin, PhD, Director

Research on instructional, communication, and information systems and their effectiveness in meeting individual and social needs.

INSTITUTE FOR RESEARCH ON ORGANIZATIONAL BEHAVIOR

Clifford P. Hahn, MS, Director

Research on human resources, selection and training, management and organization, safety, and administration of justice.

INSTITUTE FOR RESEARCH IN PSYCHOBIOLOGY

Thomas I. Myers, PhD, Director (Acting)

Studies concerned with those behavioral and physiological mechanisms which determine performance and with those factors in the environment which affect these mechanisms and thereby affect performance.

INTERNATIONAL RESEARCH INSTITUTE

Paul Spector, PhD, Director

Research on the development of human resources in developing countries; problems of working effectively abroad; evaluation of action programs in underdeveloped countries; role of attitudes and values in social change and economic development.

TRANSNATIONAL FAMILY RESEARCH INSTITUTE

Henry P. David, PhD, Director

Research on behavioral components of family interaction; coordination of transnational population research, and strengthening of family planning programs.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* American Institutes for Research 8555 Sixteenth Street Silver Spring, Maryland 20910	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

New Techniques for Measuring and Improving Reading Comprehension

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Report #1

5. AUTHOR(S) *(First name, middle initial, last name)*

Ronald P. Carver

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
15 February 1973	71	31

8a. CONTRACT OR GRANT NO. N00014-72-C-0240	9a. ORIGINATOR'S REPORT NUMBER(S) AIR -30800-2/73-TR
b. PROJECT NO. NR154-345	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Personnel and Training Research Programs, Office of Naval Research Arlington, Va. 22217

13. ABSTRACT

A standardized method has been developed which will convert prose training materials into a form which forces the trainees to read the material with at least a minimal level of comprehension. The materials, called programmed prose materials, are developed in an objective manner amenable to computer production. Phase I of the project involved an extensive investigation of a new technique, called the reading-storage test, for measuring the learning that occurs during reading so that the effectiveness of programmed prose could be properly assessed in Phases II, III, & IV. This technical report covers the Phase I and Phase II research. In six Phase I experiments, the reading-storage test was compared to two other types of tests. The results suggested that the completely objective, reading-storage test provides a better measure of the primary effects of reading than its two closest competitors, i.e., the cloze test which is developed objectively but scored subjectively, and the paraphrase test which is developed subjectively but may be scored objectively. In the Phase II experiment, programmed prose was compared to regular prose under low and high motivation conditions. The programmed prose facilitated learning under the low motivation condition, and inhibited learning under the high motivation condition, as had been hypothesized. It was concluded that: (a) the objectively developed reading-storage test is valid as a measure of the learning, understanding, comprehending, or information storing that occurs during reading, and (b) programmed prose facilitates learning in reading situations wherein attention wanes.

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Training Materials						
Prose Learning						
Reading						
Information Stored						
Reading Tests						
Cloze Tests						
Paraphrase Tests						
Reading-Storage Tests						
Reading-Input						
Programmed Prose						
Attention						

UNCLASSIFIED

Security Classification

ED 074472

AIR-30800-2/73-TR

NEW TECHNIQUES FOR MEASURING AND IMPROVING READING COMPREHENSION

Ronald P. Carver
Principal Investigator

TECHNICAL REPORT NO. 1

Prepared under Contract to the
Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research
Department of the Navy

Contract No. N00014-72-C-0240
NR No. 154-345

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for
any purpose of the United States Governement.

American Institutes for Research
Washington Office
Human Resources Research Insititute

February 1973

ERIC

ABSTRACT

A standardized method has been developed which will convert prose training materials into a form which forces the trainees to read the material with at least a minimal level of comprehension. The materials, called programmed prose materials, are developed in an objective manner amenable to computer production. Phase I of the project involved an extensive investigation of a new technique, called the reading-storage test, for measuring the learning that occurs during reading so that the effectiveness of programmed prose could be properly assessed in Phases II, III, & IV. This technical report covers the Phase I and Phase II research. In six Phase I experiments, the reading-storage test was compared to two other types of tests. The results suggested that the completely objective, reading-storage test provides a better measure of the primary effects of reading than its two closest competitors, i.e., the cloze test which is developed objectively but scored subjectively, and the paraphrase test which is developed subjectively but may be scored objectively. In the Phase II experiment, programmed prose was compared to regular prose under low and high motivation conditions. The programmed prose facilitated learning under the low motivation condition, and inhibited learning under the high motivation condition, as had been hypothesized. It was concluded that: (a) the objectively developed reading-storage test is valid as a measure of the learning, understanding, comprehending, or information storing that occurs during reading, and (b) programmed prose facilitates learning in reading situations wherein attention wanes.

ACKNOWLEDGMENTS

The data collection and data analysis were accomplished with the assistance of Debra Stein.

Pilot data was collected through the voluntary efforts of the following A.I.R. staff members: Gene Borgida, Classie Foat, Gloria Greenbaum, David Farley, Ellen Shaffer and Tania Ramashko.

Marshall Farr, Scientific Officer for the project, provided the triggering stimulus for conducting the comparison research on the paraphrase test, and Joseph Young, Task Supervisor for the project, suggested the use of the Bransford and Johnson experimental passages.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

This research project is primarily concerned with investigating the effectiveness of a method for increasing learning during reading. Another important purpose is to investigate the validity of a new method for measuring learning during reading. This introduction section will present an overview of the entire project as well as an overview of this technical report.

Previous research results have supported the theory which contends certain situations where low amounts of learning are expected, the amount learned can be increased by forcing the learner to interact or behaviorally respond to the stimulus materials. A new technique which forces an interaction with prose materials has been proposed. It is unique in that it can be objectively developed, i.e., a computer can be programmed to take prose as input and produce the training material as output. The goal of this research project is to determine the conditions under which these materials, called programmed prose materials, facilitate learning.

In order to investigate the effectiveness of the programmed prose technique, an appropriate measure of learning was needed. Previous measurement techniques, such as multiple choice questions, have usually involved subjective judgments on the part of the test developer. Subjectively developed criteria have two serious disadvantages: (a) they are costly to produce since someone must be paid a high rate for low production rates, and (b) the replicability of the results for different people and different materials is always questionable. A new method has also been developed for automatically producing standardized objective tests for amount learned from prose materials. This type of test, called a reading-storage test, has the advantage of being programmable for a computer. Given prose as input,a programmed computer can output a test for the prose material.

The entire project is to be conducted in four phases; Phases I and II were accomplished during this past year and Phases III and IV are to be accomplished during the next year. In Phase I, the validity of the reading-storage technique was evaluated. In Phase II, the effectiveness of programmed prose was evaluated in both low and high motivation conditions. Phase III will assess the effectiveness of programmed prose of varying levels of material difficulty. Phase IV will investigate the effectiveness of the method in a situation where it would be expected to be maximal, i.e., inner city high school youths reading Navy training material.

This technical report will contain the research results relevant to Phase I and Phase II, i.e., the first year of a two-year project. Before presenting the details of the research, two background rationale sections will be presented. The first section will provide the background for investigating the effectiveness of programmed prose, and the second will provide the background for investigating the validity of the reading-storage test.

Background for Programmed Prose

In 1917, Thorndike stated that the vice of the poor reader is to say the words to himself without actively making judgments concerning what they reveal. Thus, Thorndike seems to be one of the first to recognize the importance of getting the learner actively involved in the process of reading in order to facilitate learning.

More recently, this area of concern has been researched by Rothkopf (1965). Rothkopf was interested in understanding the role of student responses in programmed instruction and this prompted him to investigate specially developed adjunct questions as facilitators of learning from prose materials. Rothkopf (1970) has stated that some response on the learner's part is necessary to transform a nominal stimulus into an effective stimulus, and he has described three classes of activity involved in effecting this transformation. Class I, Orientation, involves getting Ss into the vicinity of instructional objects and keeping them there for suitable periods. Class II, Object Acquisition, involves the selection and procurement of appropriate instructional objects. Class III, Translation and Processing, involves systematic eye fixations, translation into speech, discrimination and processing. Much of Rothkopf's research has focused upon the effectiveness of questions and their placement as facilitators of learning from prose. The use of questions does not seem to be highly effective in this regard, at least when compared to an admonition to "read carefully and slowly" (see Carver, 1972a). However, Rothkopf's research does represent a concerted effort to thoroughly investigate a practical technique for facilitating learning during reading by attempting to force the learner to interact with prose materials.

Another investigator who has been concerned with the activities that the student engages in when confronted with instructional materials is Anderson (1970). The following excerpt from Anderson (1970) summarizes the nature of the problem from an instructional point of view:

> One cannot be sure what a student is doing when he is looking at the pages of a textbook. He may be reading every line or he may be skimming the page. He may test himself on the implication of what he reads, but he may not. He may give selective emphasis to certain sections as students seem to do when they underline portions of a text. The student's emphasis is not necessarily the emphasis that the teacher desires. The student may spend more time on sections that he has trouble understanding, or he may skip difficult sections. If the student gets bored or tired he may begin to daydream or even go to sleep[p.349].

Anderson points out that, traditionally, the word "attention" has been used to designate the process whereby learners translate nominal stimuli into effective stimuli. He contends also that the control of attention is probably most critical when the learners are bored, tired, influenced to work hurriedly, or given difficult material. Anderson suggests the following series of mediating processes necessary for learning: (a) noticing the stimulus, (b) translating it into internal speech, (c) evoking images for the things and events named by the words, and (d) conceiving relationships among the imagined things or events. He contends that "...the chief problem for educational engineering is to discover how to alter the characteristics of instructional tasks so as to force students to do all of the processing required for learning [p. 363]." But, Anderson also acknowledges that some procedures which force attention may in fact inhibit the complete processing necessary for understanding and learning.

Two instructional techniques have received a great deal of research attention as procedures which attempt to increase learning by forcing an interaction between the learner and prose passages. The research on the question-technique by Rothkopf has already been noted. The other technique is the cloze-technique, i.e., the deletion of certain words in some regular manner from a passage and substituting underlined blank spaces. Cloze was first recommended by Taylor (1953) as a method of measuring readability, but it has also been researched as a teaching technique (see annotated review by Jongsma, 1971).

Both the cloze-technique and the question-technique have advantages and disadvantages which are inherent to the approach, i.e., advantages and disadvantages not associated with the results they produce. The cloze-technique has the advantage that instructional materials can be developed from ordinary prose passages in a completely objective manner. The cloze-technique can be applied to any prose passage by any person (or machine) with the same general results expected no matter who or what developed the materials. Conversely, instructional materials must be subjectively developed from prose materials under the question-technique, and this is a distinct disadvantage associated with this technique. The question-technique must be applied to prose materials using the best subjective judgments of the human producers with no standards or guidelines regarding how to write the questions, the type of questions to use, the number of questions, or the location of the questions. This lack of objectivity of the question-technique not only makes it difficult to generalize research results but it also reduces its practical usefulness. An expert must be employed to produce the questions.

From the standpoint of objectivity in development, the cloze-technique has a distinct advantage over the question-technique. However, there are two primary disadvantages inherent to the cloze-technique which are not inherent to the question-technique. The cloze technique involves a mutilation of the learning materials, i.e., the prose, and this degradation of the original information should detract from learning. Some of the information is missing and all of the learners are not likely to be able to correctly fill in all of the missing parts and thereby reconstruct the original message. Thus, the cloze-technique may be expected to increase learning by forcing increased attention, but at least part of this gain is probably lost due to the fact that not all of the original information is presented. This is not a problem with the question-technique as the prose passage is always presented in its entirety; questions are added but nothing is taken away. The other primary disadvantage of the cloze-technique is that it represents a task which is quite different from ordinary reading. An individual must become a problem solver for each missing word by trying out several candidate alternatives and then choosing the one that seems best. Then, the learner must disengage himself from the primary learning task and write an answer in the blank space provided. This makes the task highly inefficient as compared to ordinary learning by reading. Since the multiple-choice task does not present this problem, it seems inherently better in this regard as compared to the cloze-technique.

It appears that both the question-technique and the cloze-technique contains serious inherent disadvantages. Recently, a method, called reading-input, has been advanced which is similar to cloze but seems to encompass its advantages while overcoming its disadvantages (Carver, 1971a). The technique has been used as a test (see Carver, 1970) and as a method for estimating material difficulty (see Carver, 1973d), but it also seemed useful as a technique to manipulate attention during learning. An example of the reading-input technique is presented in Fig. 1. Notice that it is similar to the cloze-technique except instead of deleting words and requiring fill-ins,an incorrect word is added as an alternative to the correct word.

The reading-input technique overcomes the two primary disadvantages of the cloze technique. There is very little interruption of the normal reading act since the correct answer is usually readily recognized and there is no necessity to write out a word since the simple checking of a box is all that is required. The reading-input technique also preserves the objectivity advantage of the cloze-technique; there is no subjectivity involved in the selection of the alternative wrong words. The alternative wrong words are selected from the surrounding context according to an algorithm that is amenable to computer production (see Carver, 1971b).

IN ORDER TO TAKE	COMPUTER ADVANTAGE
OF THE INFORMATION-PROCESSING	APPROACH ANALOGY
AND ITS COMPUTER CONCEPTS,	TERM A
COMPUTER MODEL OF READING	IN HAS
BEEN DEVELOPED. THE TERM	"VERY" "MODEL"
IS USED HERE TO	DENOTE AND

Fig. 1. An example of the reading-input technique applied to a segment of a passage.

The reading-input technique is very similar to programmed instruction when it is used as a method for manipulating attention by forcing the learner to interact or behaviorally respond to the prose. Thus, the reading-input technique has been termed, programmed prose, when it is used for this purpose, and the reading-input materials are called programmed prose materials.

There is nothing inherent in programmed prose materials which will allow learning to be directly manipulated. An individual who correctly chooses the alternatives in programmed prose materials will not necessarily have understood, comprehended or stored the information contained in the materials. An analogy can be made to the old saying that you can lead a horse to water but you cannot make him drink. Programmed prose materials cannot be expected to make the horse drink (the student learn) but they can be expected to make him put his mouth in the water (interact with the learning material). And, you can be reasonably sure that the horse (the learner) will never ingest any water (information) if he never puts his mouth into the water (if he never interacts).

Programmed prose would seem to be an indirect facilitator of learning in those situations where attention was not expected to be optimal. If the learner was sufficiently motivated, if the materials were not too difficult, and if the time spent learning was not too extensive, then programmed prose would probably be an inefficient way to learn. For a college student who has high ability and who is motivated to learn in his introductory psychology class, the conversion of the regular textbook prose into programmed prose is likely to inhibit the amount he learns per a fixed interval of time. This is because the lack of attention is not a problem. Programmed prose would force this individual to do things while reading that would interrupt his normally efficient learning activities. On the other hand, for students who are relatively low in ability and who are not very interested in psychology, programmed prose may force them to attend to the learning material. Therefore, the student may learn much more efficiently than otherwise would be the case. Programmed prose could be expected to facilitate learning in those situations where attention wanes, but it should inhibit learning in those situations where attention is continuously high.

The major purpose of this research project is to investigate the efficacy of the programmed prose technique in situations where it is,and is not expected to facilitate learning.

ERIC

Background for the Reading-Storage Test

Richard Anderson (1972) has contended that procedures for constructing and describing comprehension tests are a "mess" because it is impossible to know what the tests measure. He recommends that drastic action be taken. The present research on a reading-storage measure represents an attempt to solve the long recognized problem that Anderson has articulated so well. How do we measure the primary or beneficial effects of reading? What empirical measure can we use that will provide strong evidence that the individual who is given something to read has in fact comprehended, understood, or stored the information that he has supposedly read? Anderson argues extensively for the use of paraphrase questions. He contends that an individual can answer other types of questions, e.g., verbatim, using orthographic or phonological encoding without any of the comprehension that is associated with semantic encoding.

Before performing a critical analysis of the various available techniques for measuring the primary effects of reading, some background is needed for understanding the processes involved in reading. First, it is necessary to discriminate among the possible processes that an individual may engage in when prose passage in presented. The individual may scan the passage to find a particular word. The individual may rapidly skim or skip over the passage to get some idea of what the passage is about. The individual may try to memorize the passage word for word. The above activities and purposes are not exhaustive but are illustrative of the many different ways of processing the information contained in prose materials. None of the above ways are of direct interest at present. What is of interest is the type of activities normally engaged in when a person is said to be reading. These activities are considered to be a communication process wherein the thoughts of the originator of the communication (i.e., the author) are being understood, comprehended, or stored by the reader of the communication (see Skinner, 1957, p.278). Carver (1971a) has contended that the "understanding" function is automatically a type of storage function in that understanding means that the thought has been successfully related to previously stored thoughts, knowledge, or information. This "relating to" function is a type of coding function that is similar to what Pribram (1968) talks about when he discusses the basic coding of experience, i.e., Imaging. That is, most reading is an experiential process which stands as a surrogate for the following: (a) actually visiting the island that the passage describes, (b) actually participating in the happenings associated with the characters of a novel, or (c) actually witnessing the events described by a newspaper reporter. This "relating to" function is also similar to what Robinson (1960) called the fusion of ideas read with previous experience, and what Newell and Simon (1967) referred to as a chunk fixated in long-term memory or "familiarization." In a recent simulation of human long-term memory, Fridja (1972) explained the process this way:

> The same interaction between input and stored information may be used to integrate new information into the network. By linking it up to an appropriate place, available implications may become accessible and, therefore, the corresponding variety of access-ways may permit subsequent recall. Locating new information in the network may well be considered the major aspect of the process of "understanding" new input [p. 16].

One of the obvious implications of the above view of reading is that it is normally a process which involves no output or retrieval mechanisms (see Carver, 1971a). Just like any other ordinary experience, while the event is happening there is usually no subprocess which codes certain parts so that they can be retrieved later (i.e., remembered) with perfect accuracy. While experiencing or witnessing an automobile accident there is usually no subprocess which automatically goes about coding the facts that the insurance investigators will later regard as important. It is only after the event that we remember to execute these coding procedures, otherwise we may not be able to accurately retrieve those parts of the event which insurance companies regard as important. The information that is stored during reading involves a similar experiential process, i.e., no automatic retrieval codes are built into the experience so that when these codes are later cued, certain aspects of the experience can be efficiently retrieved with near perfect accuracy.

This view of the reading process, presented above, may be considered as an inefficient information-processing system. Yet, it can be argued that it is highly efficient. Unless an individual can predict quite accurately those selective aspects of his experience which will be deemed as important in the future, he will spend a great deal of his time extensively coding for retrieval many parts of his experience that will never be useful to him in the future. This would be inefficient.

Because reading is primarily an experiential event, with no automatic coding of events using specific retrieval cues, it is difficult to ascertain whether a certain stimulus event has actually been experienced or stored. If all the important aspects of a reading experience were stored in memory in certain locations identifiable by numbers, as is ordinarily the case with computers, then we could find out if a certain aspect of an event had been experienced by asking for a retrieval of the information stored using the appropriate code number. Such is not the case with reading.

If the information that is stored during reading is not stored with specific retrieval codes, how does one go about measuring the primary effects of reading? In the following excerpt, Spinelli (1970) provides a hint of how we can probe the existence of this type of stored information:

> It seems to be more economic to suggest that the basic structure of the memory system used by the brain is not addressed by location (location addressable) but by content (content addressable). What this means is that to retrieve a chunk of information all that is necessary is to provide the system with a fraction of the chunk, and the remainder will be played back [p. 295].

It appears that one of the best ways of determining whether a chunk of the information contained in a reading passage was stored is to provide a part of the original chunk and see if this will cue the remainder of the original information. Now, we are faced with how to actually put this theory into practice. The traditional use of multiple-choice questions makes use of this theory when it provides part of the original information in the question and the alternative answers. However, as Anderson (1970) has noted, the traditional use of multiple-choice questions has other large disadvantages. There are usually no guidelines or standards for: (a) what type of questions to ask or (b) how many questions to ask. Anderson (1970) has made recommendations which provide solutions to both of these problems. He states: "...in order to answer a question based on a paraphrase, a person has to have comprehended the original sentence, since a paraphrase is related to the original sentence with respect to meaning but unrelated with respect to the shape or sound of the words [p. 150]." Thus, the paraphrase suggestion provides a solution to the problem of what type of question to ask, and Anderson seems to have suggested, indirectly, a solution to the problem of how many questions to ask by mentioning the sentence as a unit. That is, a paraphrase question could be written for every sentence in a passage. Or, a proportion of the sentences could be sampled and the results generalized to the population of sentences.

The major problem with the paraphrase question approach is that the quality of the solution depends upon the artistic ability of the experimenter. As such, it has the liabilities of any subjective s 'ution. Its replicability is highly questionable and large amounts of human resources are needed for implementation. The paraphrase approach may be labeled as an objective approach since the answers to the questions may be determined objectively. Yet, the only part of this type of test that is objective is the scoring. The test development is subjective and to label such tests as objective tests is highly misleading.

Another method which has been used and recommended as a measure of the effects of reading is the cloze test (e.g., Bormuth, 1969a). Again, the primary advantage of the cloze procedure when it is used as a measure of reading comprehension is that it is highly objective from a development standpoint. Scoring of this test usually involves some subjectivity, but this is minimal. The cloze test also provides an inherent solution to the problem of what type of question to ask. Cloze also seems to provide a technique which is compatable with the theory of providing a part of the original information as a cue for playing back all of the information. In most cloze tests, four-fifths of the original information is presented and one-fifth must be played back (i.e., every fifth word is deleted).

Although the cloze test has several advantages as a measure of the primary effects of reading, it does have an inherent disadvantage that needs to be discussed. If too much of the original information is presented, the learner will process information from the test itself and be able to infer the original information in its entirety without ever encountering the original information. Anderson (1972) discusses one aspect of the problem as follows:"The trick will be to devise techniques for constructing questions that can be answered if a person has semantically encoded a communication but not answered if it has been encoded only perceptually or phonologically [p. 148.] Thus, the problem is to provide enough of a chunk of information to arouse the stored information if, in fact, it has been stored but not provide so much information that the incorrect parts on the test can be recognized from background knowledge or from the information presented in the test itself.

The cloze test, as it is normally constructed, provides so much of the original information that individuals can correctly fill in many blanks without ever storing the original information. Also, when all of the information in a passage has been stored, it is often impossible to recall the precise words used in the original passage. The above reasons explain why the cloze test has been shown to be relatively insensitive to the gain in information that accrues during reading (see Carver, 1973a, Coleman & Miller, 1968). The regular cloze test seems to provide too much of the original chunk of information to be sensitive to low degrees of information stored and the nature of the cloze task itself is such that it is not sensitive to high degrees of information stored.

It is not a simple matter to determine the quantity and quality of the original chunk of information to be provided when testing for information stored. This general problem is explained in some detail by Spinelli (1970) who states:

> If wave forms in the brain represent stimuli, responses and the consequences of responses as we have previously seen (Pribam et. al., 1967), then presentation of the stimulus will generate a playback of the whole sequence; that is to say: recognition of the stimulus, the appropriate behavior that went with the stimulus, followed by the expectation of the consequences of the behavior. The amount of extra information obtained by the network or by the organism is greater, the smaller the segment of the total input string. The amount of uncertainty, and therefore of risk for the organism in using the sequence itself becomes, on the other hand, correspondingly greater. An analogy in the auditory mode helps in understanding the significance of this parameter. The name of a song followed by the playing of the whole song will, of course, be recognized, if it has been heard before. The name of the song followed by half of the song will enable the listener to remember the remainder of the song. Ultimately, just the name of the song, or a few notes, will enable the listener to recall it entirely. But if the notes are too few, or if the name of the song is equivocal, then the level of match would be correspondingly very, very small and might not enable the recaller to identify which song we are referring to. It might be that the few notes provided are part of the beginning of many songs. Ideally, then, the acceptable match parameter should be set for that minimum value which allows unequivocal recognition of the stimulus with recall of the associate behavior and consequences of behavior.

The trick then, is to provide the minimum value of the match parameter which will trigger the retrieval of the originally stored experience. If too little of a passage is presented, then the original experience will not be reconstructed. If too much of the original passage is presented, it will be difficult to determine whether that which is played back provides evidence that the information was originally stored or whether the individual was able to correctly infer most of the original passage from the information he was presented on the test.

The original reading-storage type of test, suggested by Carver (1971a), was conceived so as to solve the above problem as well as to provide a completely objective test. However, this first suggested form of the reading-storage test was pilot tested and found to be only slightly better than the cloze test in discriminating between those who had actually read a passage prior to taking the test and those who had taken the test without every having an opportunity to read the passage. Subsequent to this finding, a number of different types of reading-storage tests were developed and pilot tested. An example of the type of reading-storage test which seemed to have the most potential is presented in Fig. 2 on the next page. The passage on which the test is based is also presented in this figure. The answers to the first few items, in Fig. 2, have been circled as an aid to understanding the following description of the test. The reading-storage (RS) test consists of the original passage in capitalized form except for every other word; only the initial letter of every other word remains. And, out of each consecutive set of five such initial letters, one has been deleted and replaced with an incorrect alternative according to an algorithm (see Carver, 1973c). The task for the individual is to read the original passage, turn to the test and without referring back to the original passage, identify the one wrong initial letter of each set of five.

The basic rationale underlying the RS test has already been presented. However, there is one additional aspect of the test that needs to be discussed. What about an idiot savant who could memorize the words without understanding or comprehending anything, and still make a perfect score on the reading-storage test; thus, erroneously indicating perfect storage, understanding, or comprehension. The frequency of occurence of such idiot savants is either zero or near zero so it does not seem necessary to be too concerned with this problem. However, the question still remains, to what extent does the simple memorization of words erroneously inflate the scores on the RS test? Although learning, for many experimental psychologists, is synonomous with memorization of words (e.g., see King, 1971), others, such as Danks (1969), have argued that the learning and comprehension process "...are not necessarily isomorphic and the variables identified as important in one situation may have a minimal effect in the other [p. 696]." The primary assumption underlying the use of the RS test that is relevant to this memorization problem concerns a normal forgetting curve. Anderson (1972) has suggested that "...a printed verbal stimulus is usually phonologically encoded and then, if it is to be remembered for more than a few moments, it is semantically encoded [p. 146]." It is assumed that memory for orthographic and phonological cues do fade quickly and that the reading-storage test would be valid only in those situations where a certain minimum body of prose was tested under certain time limit conditions. The reading-storage test would be expected to be valid

Example Passage

This is our Post Office. It is in our city. Many people work here. There is a Post Office in every city in our country. And Post Offices in every country in the world.

A Post Office helper must be honest. He must be a good worker. A Post Office helper handles lots of mail. A Post Office helper handles lots of money.

The Post Office sends letters and packages, magazines, and newspapers all over the world. It sends small animals and plants, too. It sends money for us. It saves money for us. It puts money to work for us, too.

Reading Storage Test on an Example Passage

THIS IS

1. O____ POST O____. IT I____ IN H____ CITY. M____ PEOPLE

2. W____ HERE. O____ IS A____ POST O____ IN E____ CITY

3. L____ OUR C____. AND P____ OFFICES I____ EVERY C____ IN

4. T____ WORLD. A____ POST O____ HELPER M____ BE S____. HE

5. M____ BE A____ GOOD W____. A E____ OFFICE H____ HANDLES

6. S____ OF M____. A P____ OFFICE H____ HANDLES L____ OF

7. M____. THE C____ OFFICE S____ LETTERS A____ PACKAGES, M____, AND

8. N____ ALL O____ THE W____. IT S____ SMALL W____ AND

9. H____, TOO. I____ SENDS M____ FOR U____. IT S____ MONEY

10. F____ US. A____ PUTS M____ TO W____ FOR U____, TOO.

Fig. 2. An example reading-storage (RS) test on an Example Passage.

only when an individual was given an amount of time to spend on a passage that approximated the amount required for normal reading (e.g., between about 100 and 300 words per minute). If extensive amounts of time are given, then the test is likely to become invalid as a test of the degree of understanding or comprehension because memorization of words is likely to become a primary factor. Also, when the number of words in a passage drops below 100, then it seems reasonable to become concerned about short term memory for words (i.e., orthographic and phonological encoding) as a primary factor. Certainly, the reading-storage test would not be valid for the comprehension of isolated sentences, a primary concern of Anderson (1972).

Thus, the use of the reading-storage test appears to be on a rationally sound basis when it is used to measure the understanding or comprehension which occurs during the usual processing (i.e., reading) of prose. Stated differently, there seems to be ample theoretical rationale to support the investigation of the reading-storage test as an indicator of the extent to which chunks of information, in the form of sentences, are stored during the normal reading of prose.

It is assumed that there will never be a perfect measure of comprehension, understanding, or stored experiences. This is because the concept itself is not precise enough to warrant any one perfect measure, and because better empirical indicants will only increase the probability of correctly detecting degrees of comprehension. As scores on the reading-storage test increase, the probability that comprehension occured also is assumed to increase. It does not necessarily solve the problem to make the criterion for comprehension more and more stringent, as Anderson (1970) seems to do. As the test becomes more and more stringent, the probability increases that a person who does well on the test did in fact store the information. However, while the test is made more and more stringent, the probability also increases that a person who did in fact store the information will be erroneously regarded as not having done this because he did poorly on the test. The number of errors, both Type I and Type II, need to be minimized, and it is not appropriate from a strict measurement standpoint to maximize one type of error at the expense of the other type.

The reading-storage type of measure seems to show a great deal of promise as an optimum indicator of the important and primary effects of normal reading, whether these effects be called understanding, comprehension, or stored information. One of the purposes of this research project is to investigate the validity of the reading-storage test, by studying the extent to which its purported theoretical validity can be supported with empirical evidence.

Set I Experiments

Overview

Purpose. Three separate experiments -- Experiment IA, Experiment IB, and Experiment IC -- were conducted in the Set I Experiments. The general purpose for all three experiments was to investigate the properties of the reading-storage test by comparing it to a form of the cloze test. The regular cloze test already been shown to be insensitive to the primary effects of reading (Carver, 1973a), but other forms of the test appeared to deserve full investigation.

One way to evaluate the reading-storage test is to compare the amount of gain on the test with the amount of gain as estimated from the Ss own subjective estimates of degree of understanding. Subjective estimates of the understanding of isolated sentences were used successfully by Schwartz, Sparkman, & Deese (1970), and Carver (1973a) found that such estimates were very reliable and extremely sensitive to the primary effects of reading. Carver found that these estimates approached zero when understanding would be expected to approach zero and they approached 100% when the accuracy of understanding would be expected to approach 100%. Thus, it appears that the reading-storage test, as well as the modified form of the cloze test, can be evaluated by comparing its sensitivity with the sensitivity of the understanding judgments.

Subjects. Forty-eight college students from the University of Maryland were paid to participate. The volunteers were recruited via an advertisement placed in the school newspaper. The advertisement referred to an educational research project without explaining the nature of the experiment, i.e., that it involved the administration of reading tests.

Procedure and Instructions. The Ss were tested in 14 sessions ranging in size from three to four individuals per session. The general nature of the experiment was explained at the outset. The Ss were told that they: (a) would receive $6.00 in cash at the end of the four hour experimental session, (b) could receive an additional bonus depending upon their test performance, and (c) would probably receive an average bonus around $6.00 with some individuals earning a bonus as low as $4.00 and some as high as $11.00 or $12.00.

The Ss were further informed that: (a) the testing would be conducted in three separate studies, (b) there would be a 10-minute break approximately every hour, and (c) there would be 22 short reading tests altogether.

The Ss were told that the paragraphs they would be given to read would be approximately 100 words in length and that they would be given one minute to read each paragraph prior to taking the test on the paragraph. They were told that if they finished reading a paragraph before the time limit expired, they should go back to the beginning and read it again, and keep reading until the buzzer sounded. The Ss could ascertain the amount of time remaining by observing the timer which was located immediately in front of them.

Tests. There were two types of tests -- a reading-storage type test (RS-Test) and a modified-cloze type test (MC-Test). An example of the RS-Test has already been presented in Fig. 2.

The procedures for developing the RS-Test used in this research have been described in detail elsewhere (Carver, 1973c). Briefly summarized, the procedures are: (a) retype the original passage in capital letters with 10 words per line of running text, (b) for every other word, delete all letters except the initial letter of the word, and replace the missing letters with a designating symbol such as a standard dash, and (c) for each line containing five of these skeleton type words, randomly delete one of these five initial letters and replace it with a different letter selected from the population of initial letters in the passage. The task for the S was to circle the wrong letter on each line of the test. The time limit for each RS-Test was 3 minutes. Pilot data indicated that this amount of time allowed 90 - 100% of the Ss to finish the test.

The MC-Test was a modification of the regular cloze technique. Instead of deleting every fifth word, as is usually done in the regular cloze procedure, every fourth, fifth, and sixth words were deleted and replaced with the initial letter of the word. Presented in Fig. 3 is an example of an MC-Test for the example passage in Fig.2. The task for the S was to fill in the blanks with the correct words using the initial letters as cues. As an example, the correct words have been inserted for the first one-half of the test in Fig. 3. The time limit for the MC-Tests was seven minutes. Pilot data indicated that this amount of time allowed 90 - 100% of the Ss to complete the test.

THIS IS OUR	P Post	O ffice.	I It
IS IN OUR	C city.	M Many	P people
WORK HERE. THERE	I is	A a	P Post
OFFICE IN EVERY	C ity	I in	O one
COUNTRY. AND POST	O Offices	I in	E every
COUNTRY IN THE	W world.	A a	P Post
OFFICE HELPER MUST	B be	H honest.	H He
MUST BE A	G good	W worker.	A A
POST OFFICE HELPER	H handles	L lots	O of
MAIL. A POST	O	H	H
LOTS OF MONEY.	T	P	O
SENDS LETTERS AND	P ,	M ,	A
NEWSPAPERS ALL OVER	T	W .	I
SENDS SMALL ANIMALS	A	P ,	T
IT SENDS MONEY	F	U .	I
SAVES MONEY FOR	U .	I	P
MONEY TO WORK	F	U ,	T .

Fig. 3. An example of a Modified-Cloze (MC) test for the example passage in Fig. 2.

Design. The overall design for the Set I Experiments is presented in Table 1. There were two types of experimental sessions, Type I and Type II. The Ss who participated in the Type I session were administered RS-Tests in Experiment IA and MC-Tests in Experiment IB and Experiment IC. The Ss who participated in the Type II session were administered MC-Tests in Experiment IA and RS-Tests in Experiment IB and Experiment IC. The experimental passages were all the same in all three studies, only the type of test varied from the Type I session to the Type II session.

Table 1

Overall Design for Set I Experiments

	Session	
	Type I	Type II
Experiment IA	RS-Tests (N=24)	MC-Tests (N=24)
Experiment IB	MC-Tests (N=20)	RS-Tests (N=20)
Experiment IC	MC-Tests (N=24)	RS-Tests (N=24)

Scoring. The scores on RS-Tests were percent correct scores derived from the following equation:

$$\frac{\text{Number of items right minus one-fourth the number of items wrong}}{\text{Total number of items}} \times 100 \qquad (1)$$

The percent correct score on the MC-Tests was determined by: (a) counting the number of correctly filled-in items (spelling errors were disregarded) (b) dividing by the total number of items, and (c) multiplying by 100.

Data analysis. Since many of the score distributions were highly skewed, the average value for any particular treatment-condition was estimated by the median.

Experiment IA

Purpose. The primary purpose of Experiment IA was to determine how sensitive the RS-Test was to the primary effects of reading by administering the test under both reading and non-reading conditions. The sensitivity of the RS-Test was to be evaluated by comparing it to the MC-Test and to understanding judgments. To determine the generality of its sensitivity, four different levels of passage difficulty were investigated.

Subjects. There were 24 Type I Ss and 24 Type II Ss.

Procedure. In order to familiarize the Ss with the task required by each test, they were given an example passage and an example test on the passage. The first one-half of the answers were already completed on the example test. With this example test and the original passage in front of them, the Ss were instructed to try to figure out the answers to the remainder of the test. The E gave each S individual help as was needed, and the answers were graded so as to be certain that each S understood how the test worked. Then, the Ss were informed that they would be administered a reading passage and a test on the passage for practice. All of the regular experimental procedures were employed during this practice trial.

Prior to the presentation of the practice trial, an estimation procedure was explained to the Ss. They were asked to estimate the number of complete thoughts in a passage that they understood immediately after they finished reading a passage and immediately before they started to work on the test on the passage. They were instructed that: (a) a sentence is a complete thought, and (b) the estimate could be anywhere between 0 and 100% with 0 indicating that none of the complete thoughts had been understood and 100 indicating that all of the complete thoughts had been understood. At the top of each test the following percents were typed -- 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 -- and the S was instructed to circle the one which best represented his own estimate.

The Type I Ss were informed that they would receive 5¢ for each correct answer on each test, i.e., the RS-Test, and the Type II Ss were informed that they would receive 1¢ for each correctly filled-in blank, i.e., the MC-Test. The Ss were also informed that they would be given 10 tests, but that on 5 of the 10 they would not be given the passage to read prior to taking the test.

Passages. The 100-word passages used in Experiment IA were selected from among the 330 passages studied by Bormuth (1969b). The eight passages selected were chosen to represent four levels of difficulty as indicated by the RIDE Scale (see Carver, 1973d). The RIDE Scale is simply the number of letters per word (lpw). The RIDE Levels are: Level 1, up to 4.0 lpw; Level 2, 4.1 to 4.5 lpw; Level 3, 4.6 to 5.0 lpw; Level 4, 5.1 to 5.5 lpw; Level 5, 5.6 lpw and above. Two passages were chosen to represent each of Levels 1-4. The two Level 1 passages were randomly selected from those passages with RIDE values of 3.7 and 3.8. Levels 2-4 were chosen from those with RIDE values of 4.2 & 4.3, 4.7 & 4.8, and 5.2 & 5.3, respectively. After selection, the eight passages were divided into two sets, A and B, with one passage at each level in each set, i.e., passages IA, 2A, 3A, 4A, 1B, 2B, 3B, & 4B. Two extra passages were chosen to be used as additional practice passages; one was at Level 1, 1P and the other at Level 3, 3P.

Design. Table 2 presents the design for Experiment IA. The experiment was designed to investigate the degree to which reading the passages affected the test scores at each of the four difficulty levels while controlling for: (a) differences between individuals, (b) within-level differences between passages, (c) practice, and (d) order of presentation.

Practice was controlled in two ways. First, as already noted, the Ss were administered a practice trial. Second, the first two tests, of the ten, were also regarded by the E as practice. In Table 2, it may be noted that the first two tests were exactly the same for all Ss. The Ss were paid a bonus on the basis of these tests but these data were not analyzed.

There are two primary Latin-Squares embedded in Table 1. Ss 1 - 4 were presented the A set of four tests under the reading condition (R) in one Latin-Square, and the B set under the non-reading ($\tilde{R}$) condition in another Latin-Square. Both of these four-by-four Latin-Squares were completely counter-balanced for immediate sequential effects (see Bradley, 1958). For the Ss 5 - 8, the two primary Latin-Squares were reversed so that the tests taken by Ss 1 - 4 after reading were presented without reading and the tests taken by Ss 1 - 4 without reading were presented after reading.

The within level differences between passages were controlled by using two different passages at each level for each condition -- reading and non-reading. To minimize the differences between individuals, the design in Table 2 was replicated three times, i.e., N=24. Also, the entire design was completed twice, once for the RS-Tests, Type I Sessions, and once for the MC-Tests, Type II Sessions. The order of testing alternated between 8 Type I Ss and 8 Type II Ss.

Table 2

Design for Experiment IA

	Order of Presentation									
Type of S	1	2	3	4	5	6	7	8	9	10
	$\tilde{R}$*	R*	$\tilde{R}$	R	$\tilde{R}$	R	$\tilde{R}$	R	$\tilde{R}$	R
1	1P	3P	3B	4A	1B	2A	2B	1A	4B	3A
2	1P	3P	4B	3A	3B	4A	1B	2A	2B	1A
3	1P	3P	2B	1A	4B	3A	3B	4A	1B	2A
4	1P	3P	1B	2A	2B	1A	4B	3A	3B	4A
5	1P	3P	4A	3B	2A	1B	1A	2B	3A	4B
6	1P	3P	3A	4B	4A	3B	2A	1B	1A	2B
7	1P	3P	1A	2B	3A	4B	4A	3B	2A	1B
8	1P	3P	2A	1B	1A	2B	3A	4B	4A	3B

* R is the reading condition and $\tilde{R}$ is the non-reading condition.

Order effects were controlled in two ways also. First, the two Latin Squares for the reading and non-reading conditions controlled for order within each condition. Second, the four reading tests given without reading were alternated with the four given after reading, as is indicated in Table 2 by the R and $\tilde{R}$ at the top of each column in the table.

Results and Discussion. Figure 4 contains the understanding results for both the RS-Test group and the MC-Test group. The data points in Fig. 4 represent the median, percent understanding rating for the 24 values at each difficulty level. Notice that the two curves are practically coincident suggesting between group replicability and between group comparability. One tempting generalization is that the difficult passages seem to be more difficult to understand, but this should not be inferred. Although each passage was 100 words long and the Ss were given one minute per passage, the average rate at which the passages were presented was still not equal across difficulty levels.

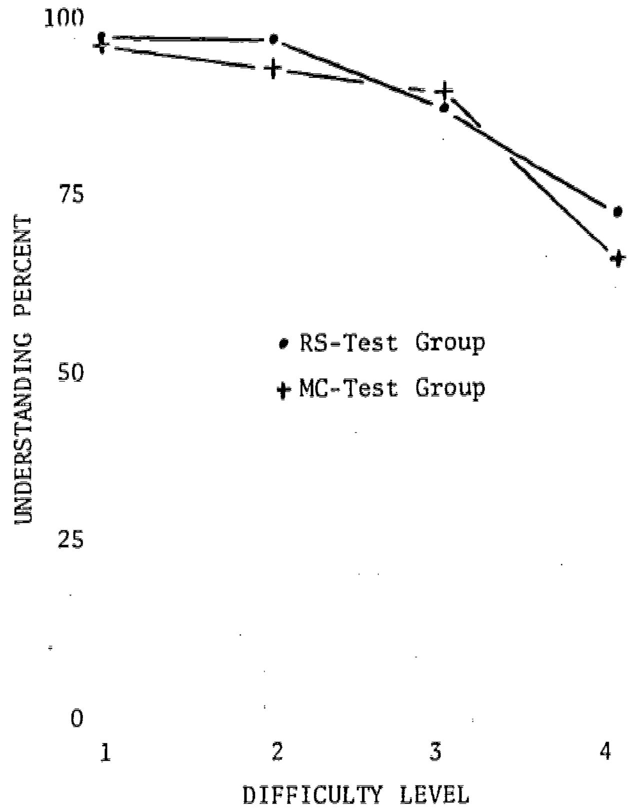


Fig. 4. Understanding as a function of difficulty for the RS-Test Group and the MC-Test Group, Experiment IA (N=48).

This is because the more difficult passages contained longer words. It has been shown that control for the physical length of the words tends to flatten out curves similar to those in Fig. 4 (see Carver, 1972b; Miller and Coleman, 1972). The Fig. 4 data can be subjected to an appropriate control procedure by calculating the efficiency of thoughts stored in standard thoughts per minute. When this control procedure is accomplished, efficiency turns out to be around 5 stpm -- 4.5, 5.3, 5.8, & 4.6 stpm -- for both groups combined at the four levels of difficulty, with no evidence of a monotonically decreasing trend. These data lend support to the theory that individuals process information from prose with the same degree of efficiency as long as the prose is not at a difficulty level which is higher than their ability level (see Carver, 1973b). The data in Fig. 4 suggest that the accuracy of understanding decreases with increases in the difficulty of the paragraphs, but when the differing passage presentation rates were controlled, it was found that the efficiency of storing thoughts was relatively equal for all four difficulty levels.

Fig. 5 presents the test scores on both the RS- and MC-Tests for both the reading and non-reading conditions. Notice that under the reading condition, the two curves are almost coincident. The MC-Test curve under the non-reading condition is somewhat erratic. Both tests seem to

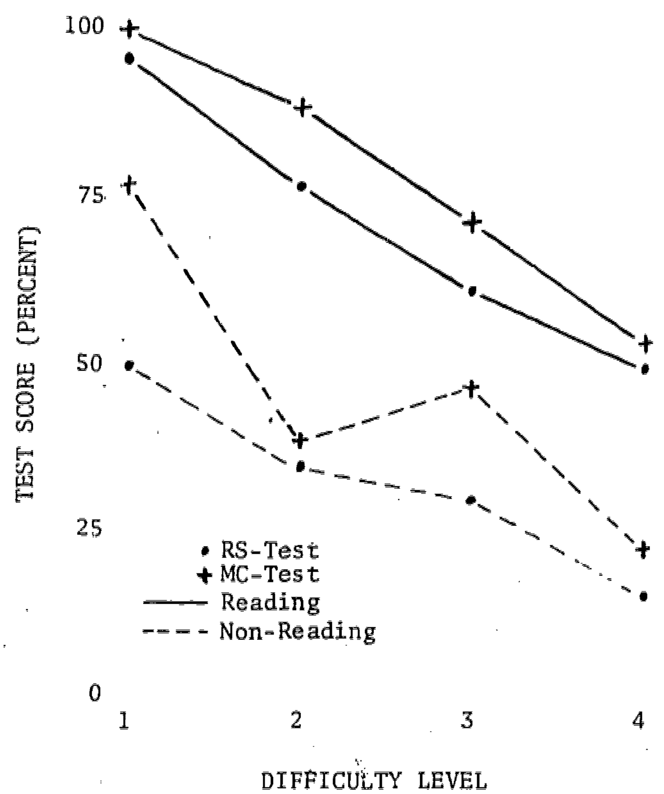


Fig. 5. Test scores as a function of difficulty level for the RS-Test and the MC-Test under reading and non-reading conditions, Experiment IA (N=48).

reflect about the same amount of average gain from the reading to the non-reading condition. Table 3 presents the gain in percentage points from the non-reading condition to the reading condition for both tests at all four levels of difficulty. The mean gain for the RS-Test was 38.2 and for the MC-Test was 31.8. The standard deviation of the MC-Test was almost twice as large,

Table 3

Percentage Point Gains and Efficiency Ratios in Experiment IA for each Difficulty level on the RS-Test and MC-Test

	Percentage Point Gain		
	Test Score	Understanding	Efficiency Ratio
RS - Test			
Level 1	45	96	.47
Level 2	42	95	.44
Level 3	32	85	.38
Level 4	34	70	.49
Mean	38.2	85.5	.45
Stand. Dev.	5.4	10.5	.04
MC - Test			
Level 1	22	94	.23
Level 2	50	91	.55
Level 3	25	87	.29
Level 4	30	63	.48
Mean	31.8	83.7	.39
Stand. Dev.	10.9	12.2	.13

10.9, as the RS-Test, 5.4. Thus, the RS-Test appears to be more sensitive and more consistent, i.e., less variable, than the MC-Test to the primary effects of reading.

Table 3 also contains the understanding gains that accompanied each test score gain. Since understanding was zero, by definition, under the non-reading condition, these values are the same as those plotted in Fig. 4. In order to assess the relative sensitivity of the objective test scores, the test score gains have been divided by the understanding gains to produce an efficiency ratio. The ratios for each level of each test are as presented in Table 3. The efficiency ratio is also the slope of the regression of test scores upon understanding, and as such is a type of validity index. If the slope, i.e., the efficiency ratio, is zero, the test scores would be considered as completely invalid since they would seem to be insensitive to the primary effects of reading. If the slope was perfect, 1.00, the test score would be considered as perfectly valid since the test was just as sensitive to the primary effects of reading as is the most sensitive indicator known.

The mean efficiency ratio for the RS-Test, .45, was slightly higher than that for the MC-Test, .39. Also, the standard deviation of the MC-Test, .13, was more than three times greater than the RS-Test, .04. Thus, this index also seems to suggest that the RS-Test is more sensitive and more reliable than the MC-Test.

These gain data in Table 3 and Fig. 5 can be directly compared to the gain data collected by Carver (1973a). These comparisons can be made because: (a) the Carver (1973a) data used paragraphs with an average RIDE Scale value of 5.05; (b) the gain associated with this 5.05 difficulty value can be found by interpolating between Level 3 (about 4.8 on the RIDE Scale) and Level 4 (about 5.3 on the RIDE Scale) in Table 3 and Fig. 5; (c) the rate of presentation of a paragraph with a RIDE Scale value of 5.06 can be estimated to be around 113 standard words per min. (swpm) since the average rate was 107.5 swpm for Level 3 and 118.2 for Level 4; and (d) the gain between non-reading and reading at 113 swpm can be found in the Carver, (1973a) data. This gain between non-reading and reading passages at a 5.06 difficulty level and 113 swpm can be calculated for all three of the measures used in the Carver (1973a) data, i.e., chunked, regular cloze, and revised cloze, as well as for the RS-Test and the MC-Test.

These percentage point gains are presented in Table 4. The highest percentage gain was associated with the chunked test. However, it should be noted that the chunked test was not developed in a standard objective manner as were the other four tests. It was developed by empirical revision procedures designed to produce items which reflect this type of gain, so it should not be surprising that it is the most effective in this regard. Of the four remaining objective test development approaches -- three forms of cloze and the RS-Test -- the RS-Test had the most percentage point gain, 33. The three cloze tests were formed by systematically deleting blanks which had to be filled in with the correct words. The traditional cloze test was formed using the every fifth word deletion pattern, and the revised cloze test was formed using an every fourth and fifth word deletion pattern. Also, the MC-Test was formed by using a fourth, fifth, and sixth word deletion pattern with the initial letter of the deleted words remaining.

Also included in Table 3 is the percentage gain in subjective estimates of understanding associated with each type of test, as well as the efficiency ratios. It should be noted that all of the understanding estimates should be equal because, theoretically, the passage difficulty and presentation rate was equal in each case. In fact, these values all are approximately equal, varying only from 70 to 77%. Notice that the chunked reading test has high validity, .80, while the RS-Test is next highest with an efficiency ratio of .43. These data suggest that the RS-Test is more than twice as valid as the regular cloze test, .19, as a measure of the primary effects of reading.

When compared to the chunked test, the RS-Test is much less sensitive to the primary effects of reading. However, this comparison is not appropriate because the RS-Test did not undergo any empirical iteration procedures designed to make it highly sensitive to gain, as did the chunked test. It is likely that an RS-Test on a particular passage could be made much more sensitive to the effects of reading if it was revised on the basis of empirical gain data. A more appropriate comparison for the RS-Test is the cloze test, because both of these types are objectively developed in a standardized manner. Also, when this comparison is made, it seems evident that the RS-Test is much more sensitive. Only the Modified Cloze test that was used in the present study appears to come close to being as sensitive to the primary effects of reading as is the RS-Test.

One result that deserves notice is the non-reading data in Fig. 5. More items can be guessed at correctly without reading on the less difficult passages

Table 4

Percentage Point Gains and Efficiency Ratios for Five Types of Tests

Type of Test	Percentage Point Gain		Efficiency Ratio
	Test Score	Understanding	
Carver (1973) Data			
Chunked	61	76	.80
Cloze	14	74	.19
Revised-Cloze	10	70	.14
Experiment IA Data			
RS-Test	33	77	.43
MC-Test	23	75	.37

as compared to the more difficult passages. Therefore, it would appear to be impossible to accurately predict the amount of information that was stored (or percent of understanding) from the absolute size of the RS-Test score. No doubt, there are also large individual differences with respect to this ability to correctly guess answers. An absolute prediction of understanding or information stored using RS-Test scores should take into account the score that would be expected without reading, given the difficulty of the paragraph and the ability of the individual.

Experiment IB.

Purpose. The purpose of Experiment IB was to investigate the properties of the RS-Test by experimentally manipulating the amount of information presented. Information was manipulated by varying the percent of the words in the reading passage which were deleted. Again, the RS-Test results were compared to the MC-Test and the understanding judgments.

Subjects. There were 20 Ss in the Type I sessions and 20 Ss in the Type II sessions.

Procedure and Instructions. The Ss were told that they would be given six tests in this experiment and that the type of test would be different from the previous experiment. They were then given the same example paragraph as in the first experiment with a different example test. If the Ss were in the Type I session,they were given the example MC-Test. If the Ss were in the Type II session, they were given the example RS-Test. Again, help was given to those individuals who had difficulty figuring out how the test worked.

The Ss were administered a practice trial using exactly the same practice passage as was used in the first experiment. This time, however, every fourth word on the passage had been omitted. The Ss were informed that they would read six passages -- one would have every sixth word omitted, one would have every fifth word omitted, one would have every third word omitted, one would have every second word omitted, and one would be administered with all words omitted (i.e., the non-reading condition).

Those Ss taking the RS-Tests were informed that they would receive 5¢ per each correct answer, and those Ss taking the MC-Tests were informed that they would receive 1¢ for each correct answer. Just as in the first experiment, the Ss were instructed to indicate their understanding estimates by circling one of the values at the top of each test.

Tests. The reading passages were taken from the five passages used in Form A of the Carver-Darby Chunked Reading Test. The five experimental passages were developed by counting the first 100 words of each passage, and then completing the passage at the end of the sentence containing the 100th word.

The RS- and MC-Tests on each passage were developed using the procedures outlined in Experiment IA.

Five different experimental conditions were developed for each reading passage. The 0% condition consisted of the passage without any of the words deleted. The 17% condition was formed by deleting every sixth word. The 33% condition was formed by deleting every third word. The 50% condition was formed by deleting every second word. And, the 100% condition was formed by deleting all words, i.e., the non-reading condition. The deletions were made by applying a white covering substance (used by typists to make manuscript corrections) on the words. The tests were reproduced by xerography thus leaving no clue as to the words omitted except for word length.

The practice passage for Form A of the Carver-Darby Chunked Reading Test was used to develop one test. This test was considered by the E as practice and was presented under only one deletion condition, every fourth word deleted (25%).

Design. Table 5 contains the design for Experiment IB. This design was employed to investigate the effect of varying passage deletions upon test scores.

Table 5

Design for Experiment IB

	Order of Presentation					
Group	1	2	3	4	5	6
1	P(25%)	1(0%)	2(33%)	3(100%)	4(17%)	5(50%)
2	P(24%)	2(17%)	3(50%)	4 (0%)	5(33%)	1(100%)
3	P(25%)	3(33%)	4(100%)	5(17%)	1(50%)	2 (0%)
4	P(25%)	4(50%)	5 (0%)	1(33%)	2(100%)	3(17%)
5	P(25%)	5(100%)	1(17%)	2(50%)	3 (0%)	4(33%)

Practice effect was controlled by regarding the first test taken by the Ss under actual experimental conditions, as additional practice. That is, the first test was always the same for all Ss, the practice passage (P) given under the 25% condition. The remaining five tests were presented in a Greco Latin-Square design to control for passage differences, order differences, and individual differences.

Each session containing a group of 3 or 4 Ss was administered tests in one of the five groups noted in Table 5. The five groups were tested in order but alternating between MC-Tests (Type I) and RS-Tests (Type II) groups. The design was completed with 20Ss (4 Ss per group) for each type of test. Since there were 24 Ss in the Type I sessions and 24 Ss in the Type II sessions, one group of four Ss of each type was an unanalyzed replication.

Results and Discussions. Fig. 6 contains the understanding results for both the RS-Test and MC-Test groups. As in Experiment IA, the data points are the median percent understanding ratings for the 20 values for each percent deletion condition.

The two curves have the same general shape, even though they are not nearly as coincident as were the understanding curves in Experiment IA. These

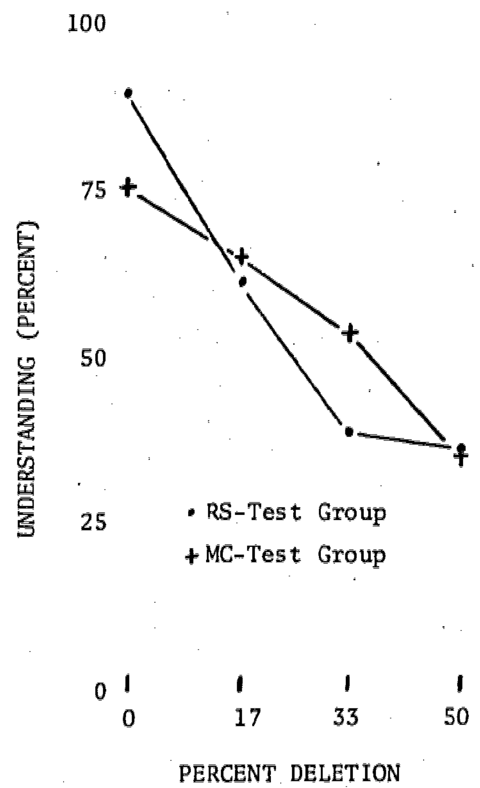


Fig. 6. Understanding as a function the percent of passage deletion for the RS-Test Group and the MC-Test Group, Experiment IB (N=40).

data do suggest reliability for the understanding variable plus comparability of RS- and MC-Test groups.

For the RS- and MC-Test groups combined, there is a decrease in percent understanding from around 81% in the 0% condition (undeleted) to around 35% in the 50% condition. The 81% understanding estimate for the undeleted condition is about the same as was obtained for the same paragraphs in the previously mentioned, Carver (1973a) study. That is, the average rate at which these five paragraphs were presented was about 121 swpm, and the understanding estimate at 121 swpm in the Carver (1973a) study can be interpolated to be about 74% for all three test groups. This result further supports the reliability of the understanding variable, and replicability of the results in both experiments.

Fig. 7 contains the percent correct scores on both the RS- and MC-Tests.

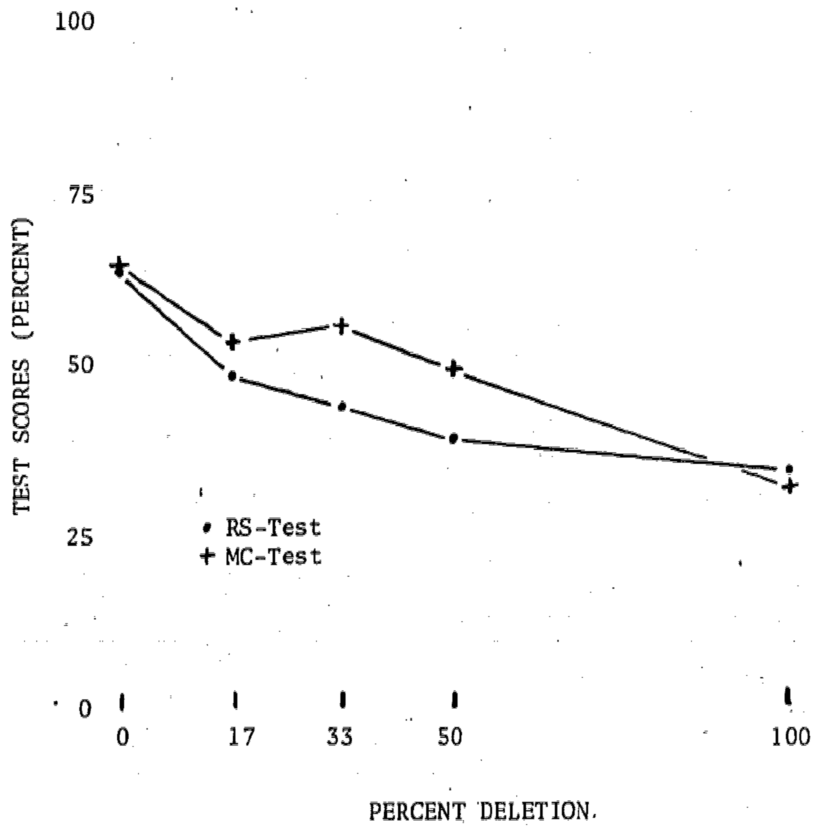


Fig. 7. Test scores as a function of the percent of passage deletion for the RS-Test and the MC-Test, Experiment IB (N=40).

Both curves have the same general shape which suggests that both measures are sensitive to the decrease in information presented. The RS-Test had a much larger decrement than the MC-Test between the 0 and 50% deletion conditions.

The efficiency ratios for the gain between the non- reading and reading conditions.

are .34 and .44 for the RS- and MC-Tests, respectively. In this experiment, the MC-Test appears to be more valid than the RS-Test. However, since both tests seemed to be measuring approximately the same thing in Experiment IA, it does not appear reasonable to interpret the small differences between these two efficiency ratios as being real nor should the differences in the two curves in Fig. 7 be interpreted as real. These differences are more likely to have resulted from uncontrolled within-individual differences. It does not seem likely that the RS-Test is the more sensitive to decrements between the 0% and the 50% deletion conditions while the MC-Test is more sensitive to decrement from the 50% to the 100% deletion condition. It seems more likely that both tests are measuring the same thing except for chance variations. The most representative curve showing the relationship between test scores and percent deletion could be found by combining the data from both the RS- and MC-Test curves in Fig. 7.

In summary, both the RS-Test and the MC-Test appear to be sensitive to decrements in the amount of information presented. The RS-Test appeared to be slightly more valid in Experiment IA and the MC-Test appeared to be slightly more valid in Experiment IB. The RS- and MC-Tests appear to be approximately equal with respect to reflecting the primary effects of reading.

Experiment IC.

Purpose. The purpose of this experiment was to investigate the effect of forgetting upon the RS-Test scores. Forgetting was manipulated by varying the number of passages and tests which were interpolated between the reading of a passage and the subsequent administration of the test on the passage.

Subjects. There were 24 Ss in the Type I sessions and 24 Ss in the Type II sessions.

Procedure. The Ss were informed at the outset that they would be administered six tests during this last experiment. The type of tests were to be exactly the same as those in the preceding experiment. They were also informed that they would be given an entire paragraph to read, as was the case in Experiment IA, but they would not be administered the tests immediately following their reading of the paragraphs. Instead, they were told that they would be given the entire set of passages to read before they started taking the tests on the passages. They were also told that the order of the tests would be reversed so that they would take the test on the last paragraph they read, first, and so on. The last test, the sixth test, would be one that would be administered without prior reading of the passage, i.e., the non-reading condition.

Again, those Ss receiving the MC-Tests (Type I session) were told that they would be given 1¢ for each correct answer, and those Ss receiving the RS-Tests (Type II session) were told that they would be given 5¢ for each correct answer.

Tests. The six, 100-word passages used in Experiment IC were developed from the six passages contained in Form B of the Carver-Darby Chunked Reading Test according to the same procedures outlined in Experiment IB. The RS- and MC-Tests were developed from the experimental passages according to the same procedures as were outlined in Experiment IB.

Design. Table 6 contains the design for Experiment IC. The 6 by 6 design for the tests is a Latin-Square completely counterbalanced for immediate sequential effects. The passage presentation order is the reverse of the test order except for the last test; instead of being given first, the passage for the last test

Table 6

Design for Experiment IC

	Passage Order					Testing Order					
Group	1	2	3	4	5	1	2	3	4	5	6
1	6	2	5	3	4	4	3	5	2	6	1
2	1	3	6	4	5	5	4	6	3	1	2
3	2	4	1	5	6	6	5	1	4	2	3
4	3	5	2	6	1	1	6	2	5	3	4
5	4	6	3	1	2	2	1	3	6	4	5
6	5	1	4	2	3	3	2	4	1	5	6

was not presented at all. Thus, the Latin-Square controls for passage differences and individual differences while reflecting the effect of interpolated activity upon test scores. Test 1 represent zero (0) interpolated activity since there was nothing interpolated between the last passage that was read and the first test which was on the last passge. Test six, 6, represents infinite (∞) interpolated activity since the passage had not been read prior to taking the test, i.e., the non-reading condition.

There were four Ss per each of the six groups in Experiment IC, N=24. The study was replicated for the MC-Tests and the RS-Tests, thus making the total of 48 Ss for Experiment IC. The six groups were tested in order, but alternating between Type I and Type II sessions.

Results and Discussion. Fig. 8 presents the RS- and MC-Test scores as a function of the degree of interpolated activity. The two curves are almost perfectly parallel. The RS-Test appears to be slightly more difficult than the MC-Test, as was also the case in both Experiments IA & IB. The MC-Test reflects

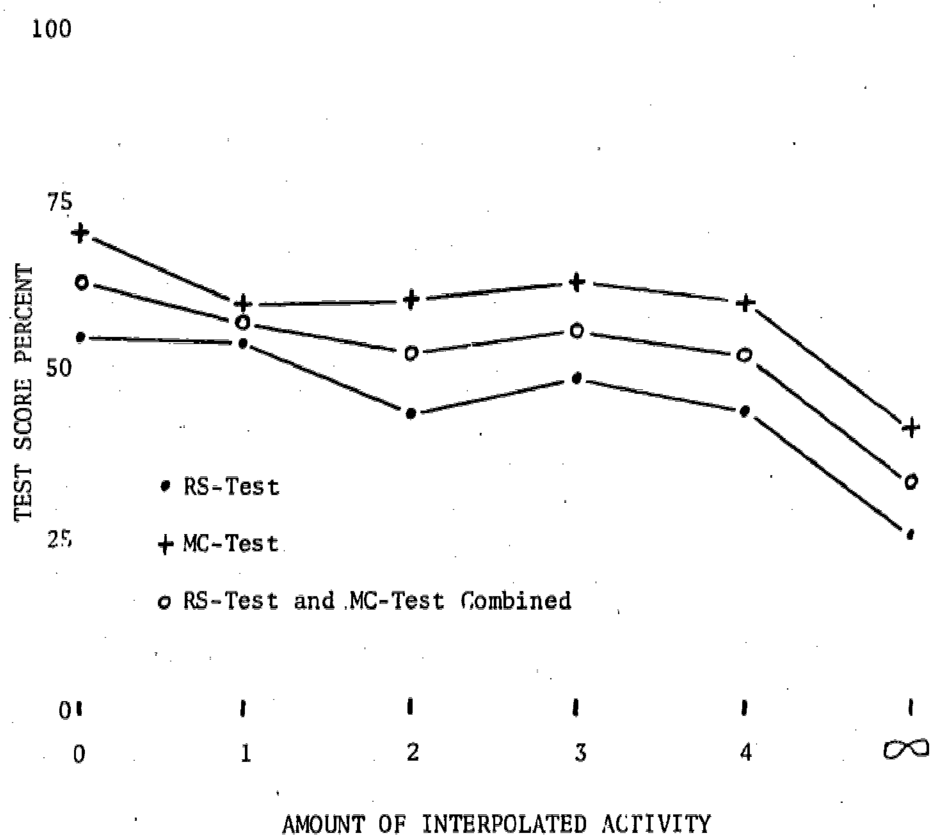


Fig. 8. Test scores as a function of the amount of interpolated activity for the RS-Test and the MC-Test and the combination of scores from both tests, Experiment IC (N=48).

more of a drop from no interpolated activity and it has a flatter plateau. Again, however, it seems more reasonable to interpret variations between these two curves as being most likely attributable to uncontrolled within-individual variations. Again, the most representative curve shape for both tests is probably the one formed by combining the data from both groups. The combined curve is also presented in Fig. 8. There appears to be a very slight drop in test scores with a small amount of interpolated activity but little drop thereafter, given the amounts of interpolated activity studied in this experiment.

It should be noted that the two curves were parallel in spite of the fact that the actual quality and quantity of interpolated activity was not the same in both groups. Both groups were given the same amount of time to read the paragraphs, but the RS-Test group received only three minutes to circle answers while the MC-Test group received seven minutes to fill in blanks with words. The actual amount of time between the reading of the first paragraph and taking the test on it was about 19 min. for the RS-Tests and about 35 min. for the MC-Tests. Since the interpolated activity was highly similar to the criterion task, there is probably not much, if any, forgetting associated with either test during at least a days time. In absolute terms, there appears to be about a drop of 8 percentage points due to interpolated activity, and this leaves a 20 percentage point residual over the score that could be expected without reading. The efficiency ratio can be used to interpret the amount of forgetting that occurs with interpolated activity. That is, an efficiency ratio of about .40 can be used to convert a test score loss of 8 percentage points into a loss of 20% in accuracy of standard thoughts stored. Therefore, it may be said that these Ss experienced about a 20% loss in information stored due to the forgetting that occurs with interpolated activity (or the passage time) but they still remembered about 50% (20/.40) of the original information after a great deal of interpolated activity.

These data can be interpreted as favorable for the use of the RS-Test. If the scores had dropped sharply with any interpolated activity or had dropped to that of the non-reading condition with these amounts of interpolated activity, then the data would not seem to have been reflecting information stored. Instead, the RS-Test would have been reflecting an immediate, or short term, memory for words. These data suggest that the RS-Test is indeed reflecting the type of information that is stored during normal reading because this type of information does seem to fade slightly immediately after it is stored, but it certainly does not fade to anywhere near zero within the same day that it was stored.

Conclusions and Implications.

From these data it may be concluded that the reading-storage type of test is valid as an indicator of the primary effects of reading. None of the various cloze techniques seem to be more valid than the reading-storage test while some, including the regular cloze procedure, appear to be considerably less valid. The modified cloze test seems to be more valid than other varieties of the cloze procedure, especially the regular cloze test, and it appears to be about as valid as the reading-storage test. Therefore, the reading-storage test appears to be at least as valid as other objectively developed types of test. Since cloze tests are not 100% objective, i.e., they must be scored subjectively, the reading-storage test would seem to be preferred over cloze tests as a measure of the primary effects of reading.

Two questions about the validity of the reading-storage test remain to be explored. First, how does it compare to the highly recommended, but subjectively developed paraphrase test? Second, just how much is the reading-storage test influenced by the short-term memorization of words? The forgetting curve data in Experiment IC suggested that the reading-storage test was sensitive to the primary effects of reading as opposed to the memorization of words, but more and different evidence is needed here. The next set of experiments will provide further evidence relevant to these aspects of the validity of the reading-storage test.

Set II Experiments

Overview.

Purpose. There were three primary purposes in the Set II Experiments: (a) to determine the extent to which the RS-Test is influenced by the memorization of words, (b) to compare the RS-Test to the paraphrase test, and (c) to determine the effect of programmed prose upon learning from prose materials under low and high motivation conditions.

There were four experiments relevant to these purposes. One experiment was relevant to the first purpose. All four experiments were relevant to the second purpose. There was also one experiment relevant to the third purpose.

Subjects. Fifty-eight college students were paid to participate. Again, as in the Set I Experiments, the volunteers were recruited from the University of Maryland via an advertisement in the school newspaper. Again, the advertisement referred to an educational research project without explaining the nature of the experiment, i.e., that it involved the administration of reading tests.

Procedure and Instructions. There were five experimental sessions with the size of each group ranging between 11 and 12 individuals each. At the outset, the Ss were told that: (a) they would be paid $10.00 in cash at the end of their four hours of participation, (b) there would be three 10-min. breaks during the afternoon, (c) they would be taking a number of short reading tests, and (d) their scores on the tests would be mailed to them. After these general instructions, a set of standardized tests were administered followed by Experiment IIA, Experiment IIB, Experiment IIC & Experiment IID.

Standardized Tests. All 58 Ss were administered two standardized tests, the Basic Reading Rate Scale (BRRS) and the Reading Level 4 (RL-4) test. These two tests were administered primarily to provide a control over individual differences in Experiment IIb. The BRRS is a published test which measures the rate at which easy prose can be read under conditions which control for comprehension. It provides for the categorization of readers into four types -- Beginning Readers, Good Readers, Better Readers, and Best Readers.

The RL-4 is an unpublished test which was developed by applying the reading-input technique to five, 100-word paragraphs at the college level of difficulty,

ERIC

i.e., at RIDE Level 4. The Ss were instructed to try to get as many right as possible in as short a length of time as possible. The primary score on this test was an efficiency score, i.e., number correct (corrected for guessing) divided by the time required in min.

Scoring and Data Analysis. The scoring and data analysis in the Set II Experiments were accomplished in exactly the same manner as in the Set I Experiments except for the programmed prose experiment, Experiment IIB. The scoring and data analysis for Experiment IIB is explained in detail later. The scoring of the paraphrase test was performed in the same manner as the RS-Test.

Experiment IIA.

Introduction. The primary purpose of this experiment was to investigate the extent to which the RS-Test reflects the primary effects of reading as opposed to simply the memorization of words.

In order to achieve this purpose, it seemed desirable to be able to manipulate information stored, understanding, or comprehension under conditions wherein memorization for words should remain constant. Bransford and Johnson (1972) have recently developed two passages which seem to permit this type of experimental manipulation. When these passages are administered with context cues, high comprehension supposedly results, but when these passages are administered without context cues little or no comprehension results. For example, when a picture is viewed briefly, prior to reading, individuals seem to have no trouble in comprehending what they read. However, individuals who are given the same passage without the opportunity to view the picture seem to comprehend little of what they read.

One way to test the validity of the RS-Test would be to administer it under both conditions, i.e., with and without context cues. If the RS-Test is primarily reflecting the ability of individuals to memorize words, then there should be little or no difference between the RS-Test scores under these two conditions since the opportunity for memorization of words is equal. If the RS-Test is primarily reflecting the normal effects of reading, then there should be a large gain when the context cue is given.

Experiment IIA used these two specially developed passages to investigate the sensitivity of the RS-Test to the primary effects of reading. Furthermore, paraphrase test (P-Test) questions also were developed on these same passages so that the validity of the RS-Test could be compared to the validity of the P-Test.

Subjects. There were a total of 48 Ss in this experiment. These Ss were the first 48 of the total of 58 who participated in the Set II Experiments.

Procedure and Instructions. The Ss were given instructions, examples, and practice on the: (a) RS-Test, (b) the P-Test, and (c) the understanding judgments. These procedures were the same as those employed in the Set I Experiments except that in this experiment all the Ss learned how to take both types of tests before they were actually given either type of test. Also, the Ss in this experiment did not have the opportunity to practice taking a test under the non-reading condition.

Passages. Passage A had a picture as a context cue. The passage contained 132 words and 9 sentences, and was at RIDE Level 3 difficulty, 4.6 letters per word. Passage B had a two word title for a context cue. The passage originally contained 181 words and 15 sentences. However, to make it comparable in length to Passage A, the last 4 sentences were deleted. Thus, Passage B contained 136 words and 11 sentences, and was at RIDE Level 2 difficulty,4.4 letters per word. These two paragraphs actually had very close difficulty estimates, 4.6 and 4.4, with Passage A being at the bottom of Level 3 and Passage B being at the top of Level 2.

Tests. The RS-Tests were developed for the two passages according to the standard algorithum discussed in the Set I experiments.

The P-Test questions were developed according to the recommendations given by Anderson (1972). Anderson defined two statements as being paraphrases of one another "... if (a) they have no substantive words (nouns, verbs, modifiers) in common and (b) they are equivalent in meaning [p. 150]." Anderson then states that to form a test item from a paraphrase, "... you delete an element, which is to be supplied or identified by the student, or you transform a segment of the paraphrased statement into a question [p. 151]." These rules of Anderson's were followed closely in the initial stages of item development but insurmountable difficulties developed. The level of redundancy between sentences in a passage is such that E found it to be exceedingly difficult (approaching impossible) to develop a paraphrase item for each sentence which E thought S could not get correct without ever reading. It was suspected that Anderson had used his rules only for isolated sentences instead of prose passages. The paraphrase guidelines of Anderson were modified by writing 5-choice, multiple-choice items on each sentence. The question and correct alternative represented a paraphrase of the original sentence whereas the incorrect alternatives resulted in meaning changes from the original sentence.

Design. There were two different passage conditions (Passage A and Passage B), and there were two different test type conditions (RS-Test and P-Test). There were three different reading conditions -- Context, No Context, and Non-Reading. In the Context condition, either the picture (for Passage A) or the title (for Passage B) was given prior to reading the passage. In the No Context condition, neither of the two context cues were given prior to reading. In the Non-Reading condition, the test on the passage was given with no opportunity to see the context cue or read the passage. Finally, there were two different orders of presentation, first or second, and this constituted the fourth factor that was manipulated.

There were total of 24 different treatments (2 x 2 x 3 x 2). The first 24 Ss to participate in the five sessions were administered each of the 24 treatments, and the second group of 24 Ss were administered each of the 24 treatments again. The final 10 Ss were also administered these treatments, but their data were not analyzed.

Time Limits and Control Pages. The passages and tests were assembled into a booklet with a cover page. The Ss were given 15 sec. to look at the first page in their booklet. For the Ss in the Context condition, this was either a picture or a title. For the Ss in the No Context and Non-Reading condition this was a control page containing brief instructions to the Ss to keep their eyes on their own tests and not to turn the page until they were told to do so.

The Ss were given one min. to look at the next page. For the Ss in the Context and No-Context conditions, this page contained either Passage A or B. For the Ss in the Non-Reading condition, this page contained directions to simply sit quietly until the time limit was up, since they would not have the opportunity to read the passage before taking the test on it.

The test followed the second page. The Ss were given three minutes to work on the test and this seemed to be ample time for about 90 - 100% of the Ss to finish.

After the test on the first passage was completed, the 15 sec., 1 min., and 3 min. time limits were repeated for the second passage treatments.

Results and Discussion. Fig. 9 contains a plot of the median values for the understanding data under the No Context and Context conditions. All of the

Ss had been instructed to circle a zero understanding estimate under the Non-Reading condition so these data have not been plotted. Each data point in Fig. 9 represents eight Ss. There is a curve for each test-passage treatment, i.e., RS-Test A, RS-Test B, P-Test A, and P-Test B. Notice that there is an increase

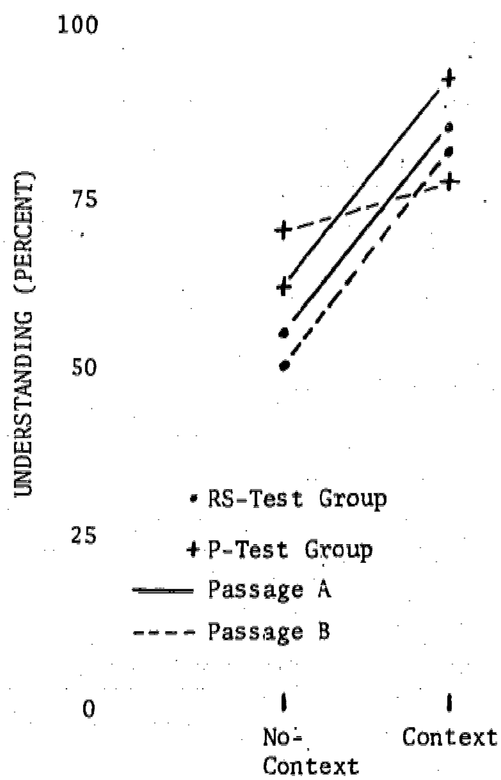


Fig. 9. Understanding as a function of the context condition for Passage A and Passage B as estimated by individuals in the RS-Test Group and the P-Test Group, Experiment IIA (N=48).

of about 30 percentage points in the understanding ratings for three of the four treatments. There was little or no gain associated with the P-Test B treatment, i.e., 7 percentage points. These data suggest that there was something different about the individuals in each of the two groups representing the two data points for P-Test B. This rating was made prior to taking the test so that there should be no difference, except for the uncontrolled individual differences, between this curve and the RS-Test B curve. Taken collectively, the four lines in Fig. 9 support the findings of Bransford and Johnson (1972) which indicate that context cues are important for comprehension or understanding.

Fig. 10 contains a plot of the median values for both tests on both passages under all three reading conditions. Each data point in Fig. 10 also represents eight Ss. It may be noted that the gain due to context, i.e.,

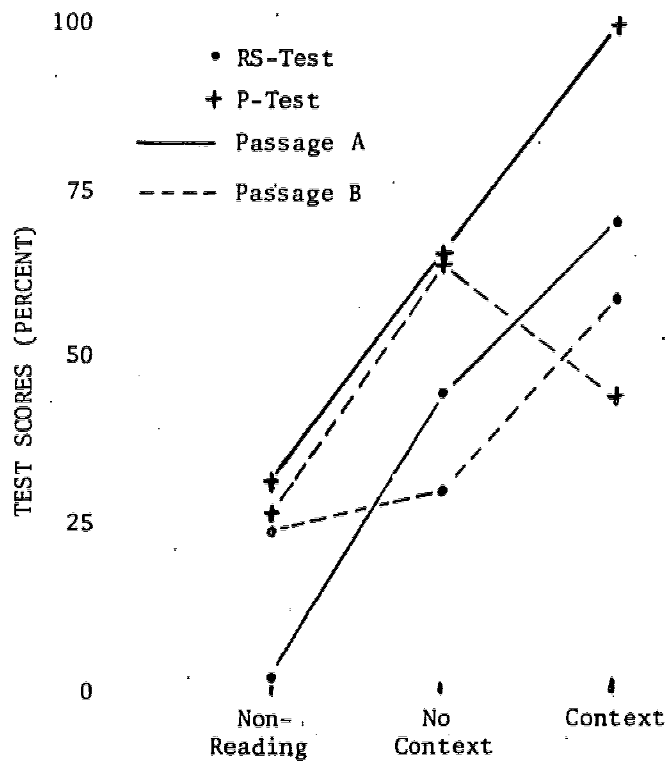


Fig. 10. Test Scores as a function of reading condition for Passage A and Passage B as measured by both the RS-Test and the P-Test, Experiment IIA (N=48).

from the No Context to the Context conditions, was substantial for the RS-Test, i.e., 26 and 29 percentage points for Passages A and B, respectively. This finding is even more impressive when it is compared to the corresponding percentage gains for the P-Test, 34 and -20, respectively. The -20 gain for the P-Test B condition should be discounted because the understanding data in Fig. A indicates that the comprehension of the two groups involved in this gain score was not equivalent to the other groups prior to taking the tests. These data suggest that the RS-Test does reflect the primary effects of reading and reflects these effects just as well if not better than the paraphrase type of test.

The gain from the Non-Reading to the No Context condition might have been interpreted as being primarily due to memorization were it not for the fact that the gain on the P-Test was so substantial. Both the P-Test A and the P-Test B reflected approximately the same percentage point gains from the Non-Reading to the No Context condition, i.e., 34 and 37, respectively. The corresponding gains for the RS-Test were 42 and 6, respectively. The average gain for the P-Test was 35.5 and the corresponding value for the RS-Test is 24.0. If the gain from Non-Reading to No Context is primarily due to memorization, then the paraphrase type of test seems to be more influenced by memorization than does the reading-storage type of test.

The preceding analysis and interpretation of the data is summarized by the efficiency ratios presented in Table 7. These data represent the average test score gains and understanding gains for the two passages (A and B) combined.

Table 7

Percentage Point Gains and Efficiency Ratios for Experiment IIA

	Percentage Point Gain		
	Test Score	Understanding	Efficiency Ratio
RS-Test			
Non-reading to No Context	24.0	52.5	.46
Non-Reading to Context	51.5	83.5	.62
P-Test			
Non-reading to No Context	35.5	66.0	.54
Non-reading to Context	42.5	84.5	.50

Notice that the efficiency ratio for the Context condition was higher for the RS-Test, .62, than it was for the P-Test, .50, and the efficiency ratio for the No Context condition was lower for the RS-Test, 46, than it was for the P-Test, .54. Thus, these efficiency ratios also suggest that the RS-Test is more sensitive to the primary effects of reading than is the P-Test, and it is less sensitive to the effects of word memorization than is the P-Test.

The evidence from this experiment further supports the validity of the RS-Test as a measure of the primary effects of reading. There is no evidence that scores on the RS-Tests are influenced significantly by the memorization of words. It seems to be just as sensitive to the primary effects of reading as is the paraphrase test.

The experiment which followed this experiment was Experiment IIB. However, Experiment IIB will not be presented next because its primary purpose was not to compare the RS- and P-Tests. Experiment IIC and Experiment IID, which follows, were designed to compare further the RS-Test and the P-Test.

Experiment IIC.

Introduction. The primary purpose of this experiment was to investigate the sensitivity of the RS-Test to the primary effects of reading as compared to the P-Test. Again, as in Experiment IA,the differences between reading and non-reading paragraphs at four different levels of difficulty were investigated.

For the RS-Test, this experiment was almost an exact replication of Experiment IA. This time, however, an additional dependent variable was added. The passages were typed in all capital letters with sentence ending punctuation and spacing cues omitted, and the S's task was to indicate where he thought the sentences ended by placing an X mark there, as he read the passage. If the S could not mark correctly the sentence ending punctuation, then he could not be expected to have stored the information. This task provided objective evidence regarding the comparability of the individuals after reading and before they took either the RS- or P-Test.

Subjects. As in Experiment IA, there were a total of 48 Ss who participated in this experiment. These Ss were the first 48 of the total of 58 who participated in the Set II Experiments.

Procedures and Instructions. The procedures and instructions were exactly the same as in Experiment IA with the following exceptions: (a) the Ss received instructions and practice regarding the placement of an X mark between sentences as they read the passages; and (b) each S was administered both types of tests without knowing which type would follow while he was reading a particular passage.

During the experiment which preceded this experiment, i.e., Experiment IIB, the Ss were informed that they would receive bonuses for their performance during the remainder of the session. For this experiment, Experiment IIC, they were told that they would receive 2¢ for each sentence they marked correctly. This scoring system was explained to them as follows: "If you fail to mark the end of the sentence or, if you place an extra X mark somewhere else in the sentence, the sentence will be scored as incorrect and you will not receive 2¢ for it." The Ss were also told that they would receive 2¢ for each answer they marked correctly on the 10 tests.

Passages. The passages were exactly the same as in Experiment IA except they had been retyped entirely in capital letters with sentence ending punctuation and spacing cues omitted.

Tests. The RS-Tests were exactly the same tests as were used in Experiment IA.

The P-Tests were developed by E on each passage using the same item development guidelines as were described in Experiment IIA.

Design. The design of Experiment IIC was exactly the same as Experiment IA, except for one difference. Instead of each S receiving only one type of test, i.e., an RS-Test group and a P-Test group, each S received both types of tests according to the design presented in Table 8.

Table 8.

Design for Experiment IIC

	Order									
	1P	3P	1	2	3	4	5	6	7	8
Group	$\tilde{R}$*	R*	$\tilde{R}$	R	$\tilde{R}$	R	$\tilde{R}$	R	$\tilde{R}$	R
A	RS**	P**	P	RS	RS	RS	P	P	RS	P
B	P	RS	RS	P	P	P	RS	RS	P	RS

* R is the reading condition and $\tilde{R}$ is the non-reading condition.
** RS signifies the RS-Test and P signifies the P-Test.

Again, there were 10 tests altogether but the first two, 1P and 3P, were exactly the same for all Ss in each of Groups A and B and they were regarded by the E as practice. The remaining 8 tests were varied according to the same Latin-Square design as was used in Experiment IA except that one group of Ss, Group A, received a different type of test on each passage at each corresponding order position as compared to the other group, Group B. The R and $\tilde{R}$ in Table 8 signifies the reading and non-reading conditions respectively.

The first 8 Ss tested were placed in Group A, the second 8 Ss tested were placed in Group B, and this alternating between Group A and Group B continued for each consecutive set of 8 Ss until the data from all 48 Ss had been collected.

Results and Discussion. Fig. 11 contains the median percent of the sentences marked correctly at each level of difficulty for those Ss who were subsequently administered RS-Tests and P-Tests. If the two groups and two sets

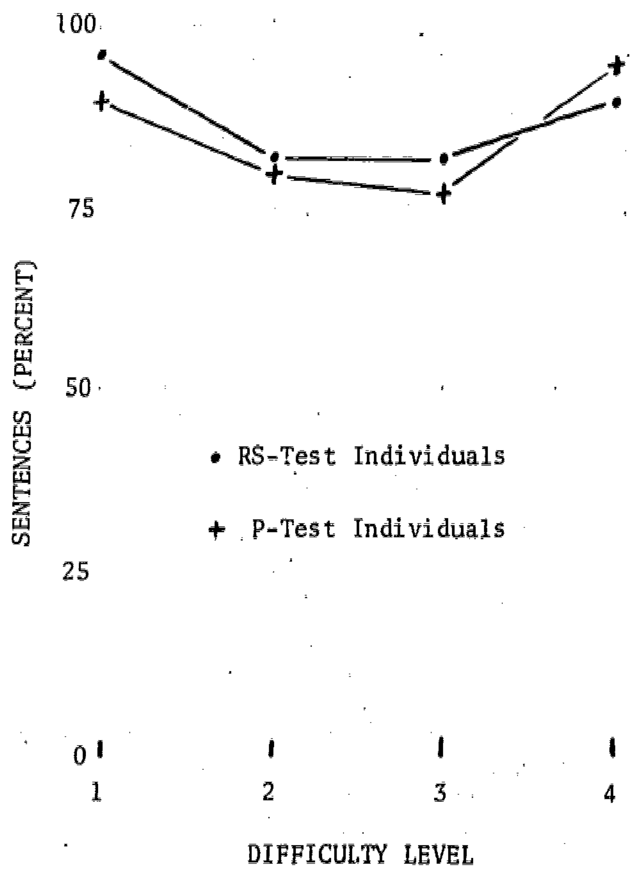


Fig. 11. Percent of sentences marked correctly as a function of difficulty level for those individuals who were administered RS-Tests and P-Tests, Experiment IIC (N=48).

of conditions were equivalent, then these curves would be perfectly coincident, and, the two curves were almost perfectly coincident. For the first three levels of difficulty the individuals who took the RS-Tests scored slightly higher than those who took the P-Tests. At the highest level of difficulty, Level 4, the RS-Test individuals scored slightly lower than those individuals who took the P-Tests. These results indicate that around 80 - 90% of the sentences could be determined correctly by the Ss, no matter what the difficulty level.

Fig. 12 presents the median percent understanding estimates at each level of difficulty for those who took the RS-Tests and those who took the P-Tests.

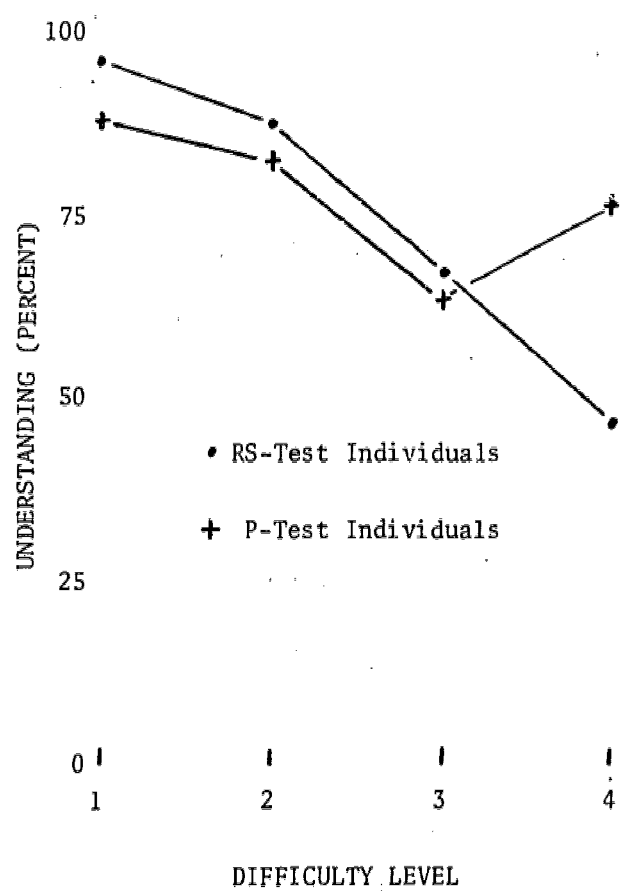


Fig. 12. Understanding as a function of difficulty level for those individuals who were administered RS-Tests and P-Tests, Experiment IIC (N=48).

The RS-Test individuals reported almost the same degree of understanding on the first three levels, but at Level 4 the P-Test individuals reported a somewhat higher understanding estimate. These data parallel the sentence data in Fig. 11, except the interaction at Level 4 is somewhat greater for the understanding judgments than it was for the sentence marking task. These data in Fig. 12 essentially replicate the corresponding data from Experiment IA presented in Fig. 4. Understanding in this experiment seems to be slightly below that of Experiment IA, as would be expected from the fact that in Experiment IIC the Ss had to figure out where the sentences ended while they were reading.

The data from Fig. 11 and Fig. 12 indicate that after reading and prior to taking either the RS- or P-Tests, the individuals at Levels 1 - 3 were almost exactly equivalent with respect to storing the information contained in the passages. The RS-Test individuals seemed to have understood slightly more than the P-Test individuals for Levels 1 - 3. At Level 4 the P-Test individuals seemed to have understood more.

Fig. 13 contains the RS-Test and P-Test scores for the four levels of difficulty and under both the reading and non-reading conditions. Under the reading condition, the two curves are roughly parallel. Their largest difference is at Level 4 where the P-Test individuals scored much higher than the RS-Test individuals, but these data are in keeping with the differences found earlier in sentence marking (Fig.11) and understanding estimates (Fig. 12). Under the non-reading condition, the P-Test scores are highly erratic from level to level. About 44% of the Level 2 questions could be answered without ever reading the passages while about 0% of the Level 4 questions could be answered without reading them. The RS-Test appears to present the most consistent results, although on the average the P-Test appears to reflect the most gain due to reading. This impression from Fig. 13 is supported by the gain data for the four levels presented in Table 9. The average gain for the P-Test is 55.0 percentage points and the average gain for the RS-Test is 37.0 percentage points. Therefore, the P-Test seems to be slightly more valid, on the average, than the RS-Test for reflecting the primary effects of reading, in this experiment. However, the standard deviation of the P-Test gains, 15.1, was almost twice as large as the RS-Test, 9.7. Thus, the RS-Test appears to be more consistent, i.e., reliable. This interpretation also is supported by the efficiency ratio data in Table 9.

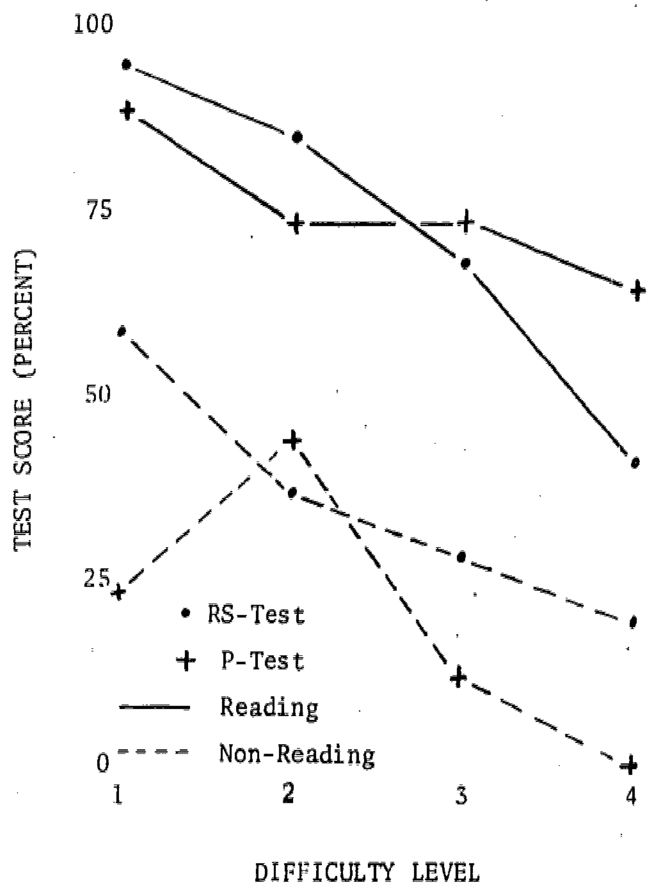


Fig. 13. Test score as a function of difficulty level for the RS-Test and the P-Test under both reading and non-reading conditions, Experiment IIC (N=48).

The average efficiency ratio was higher for the P-Test, .73, as compared to the RS-Test, .51. However, the standard deviation of the four P-Test efficiency ratios, .24, was three times greater than the standard deviation of the four RS-Test efficiency ratios, .08.

These data may also be compared to the data in Table 4, i.e., the efficiency ratios for a hypothetical passage having a difficulty index of 5.06 being read at 113 swpw. In this experiment, these efficiency ratios were .54 and .91 for the RS- and P-Tests, respectively. The efficiency ratio for the RS test, .54, was somewhat higher than it was in Experiment IA, .43. The efficiency ratio for the P-Test, .91, was even higher than the chunked test, .80, and it approached the ideal value, 1.00.

These data reflect the item writing ability of the E for the P-Test as well as the nature of the two tests. It appears that E did not do a good job of writing the Level 2 items. However, since the E used the same techniques as guidelines for all tests, these data also reflect the unreliability of this subjective method.

The RS-Test results for the non-reading condition represent an almost perfect replication of the Experiment IA data. These data are presented in Fig. 14. This result was expected because these conditions were exactly the

Table 9

Percentage Point Gains and Efficiency Ratios in Experiment IIC for each Difficulty Level on the RS-Tests and P-Tests

	Percentage Point Gain		
	Test Score	Understanding	Efficiency Ratio
RS-Test			
Level 1	37	96	.39
Level 2	49	87	.56
Level 3	40	67	.60
Level 4	22	47	.47
Mean	37.0	74.3	.51
Stand. Dev.	9.7	18.9	.08
P-Test			
Level 1	64	87	.74
Level 2	29	82	.35
Level 3	61	63	.97
Level 4	66	76	.87
Mean	55.0	77.0	.73
Stand. Dev.	15.1	9.0	.24

same between the two experiments. These data indicate one dimension of passage difficulty, i.e., as passages become less difficult, as measured by the RIDE Scale, more items can be answered correctly without reading.

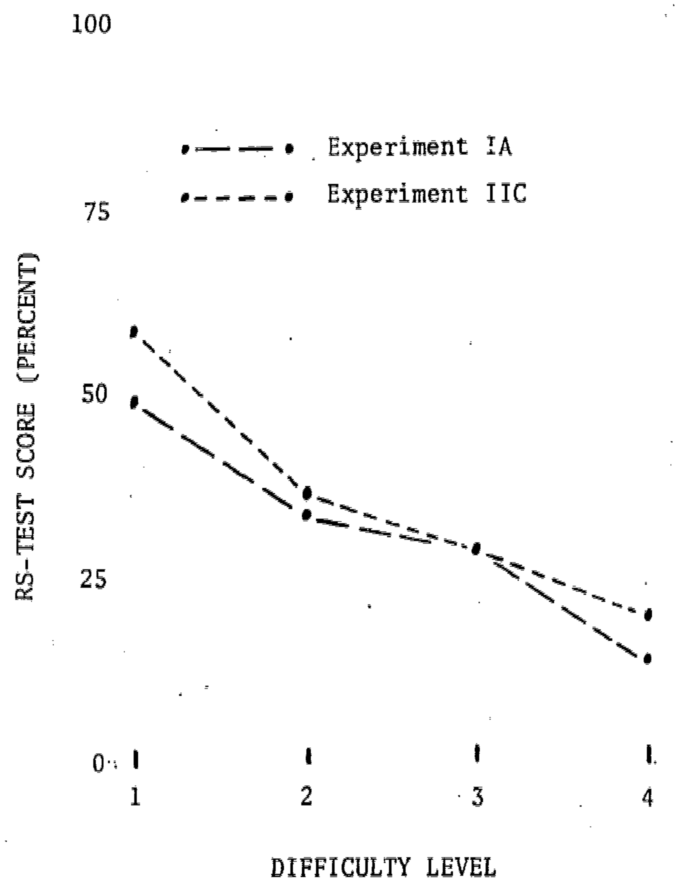


Fig. 14. RS-Test score as a function of difficulty level under the non-reading condition for Experiment IA(N=48) and Experiment IIC (N=48).

This relationship appears to be approximately linear.

Conclusion.

From this experiment it would appear that the P-Test is, on the average, slightly more sensitive than the RS-Test to the primary effects of reading but the RS-Test appears to be more reliable than the P-Test.

Experiment IID.

Introduction. The purpose of this experiment was to compare the RS-Test and the P-Test under conditions wherein the information stored from a passage was directly manipulated. This experiment was a replication of Experiment IB with the following exceptions: (a) Instead of comparing the MC-Test with the RS-Test, the P-Test was compared to the RS-Test, and (b) the reading passages had been typed in capital letters with sentences ending punctuation and spacing omitted so that the Ss had to mark the sentences as they read.

Subjects. There were 40 Ss who participated in this experiment. These Ss were the first 40 of the total of 58 who participated in the Set II Experiments.

Procedures and Instructions. The procedures and instructions were exactly the same as in Experiment IB with the exception that both the RS-Test Group and the P-Test group were informed at the outset that they would mark sentences as they read passages in a manner similar to what they had done in the preceding experiment, i.e., Experiment IIC.

Passages. The passages were exactly the same as in Experiment IB except they had been retyped entirely in capital letters with sentence ending punctuation and spacing cues omitted. The deleted words were replaced with standard length dashes so that the word length cue was not available in Experiment IID as it was in Experiment IB.

Tests. The RS-Tests were exactly the same tests as were used in Experiment IB. The P-Tests were developed by the E on each passage using the same item development guidelines as were described in Experiment IIA.

Design. The design of this experiment was exactly the same as Experiment IB except the P-Test replaced the MC-tests.

Results and Discussion. Fig. 15 contains the median percent of sentences answered correctly as a function of the percent of the passage that had been deleted. For the RS-Test group there was a 63% decrease in the percent of sentences answered correctly between the no-deletion (0% condition) and the every other word deletion (50% condition), and this decrease was almost perfectly linear. Also,this result was almost perfectly replicated by the P-Test Group, 56% decrease. These data indicate that information stored was being directly manipulated by the

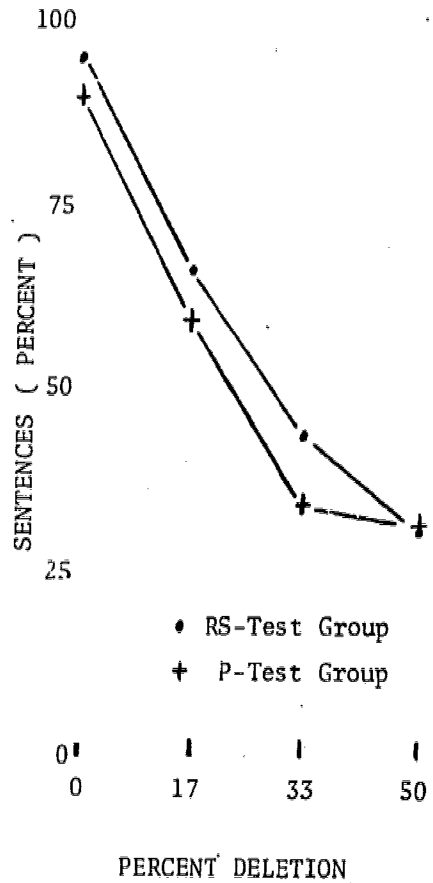


Fig. 15. Percent of sentences marked correctly as a function of the percent of passage deletion for the RS-Test Group and the P-Test Group, Experiment IID (N=40).

deletion treatment, and that both the RS- and P-Test groups were approximately equal prior to being administered the test.

Fig. 16 presents the median understanding percents as a function of the degree of deletion. There is a 49% drop in percentage points for the RS-Test Group between the 0% condition and the 50% condition. There is a corresponding 50% drop in the P-Test, and the P-Test group data are almost perfectly parallel to the RS-Test group. However, for these data the P-Test group reports consistently higher understanding ratings, about 15% higher. It appears that there may be somewhat of a "halo" type of effect for the understanding estimates on the P-Test.

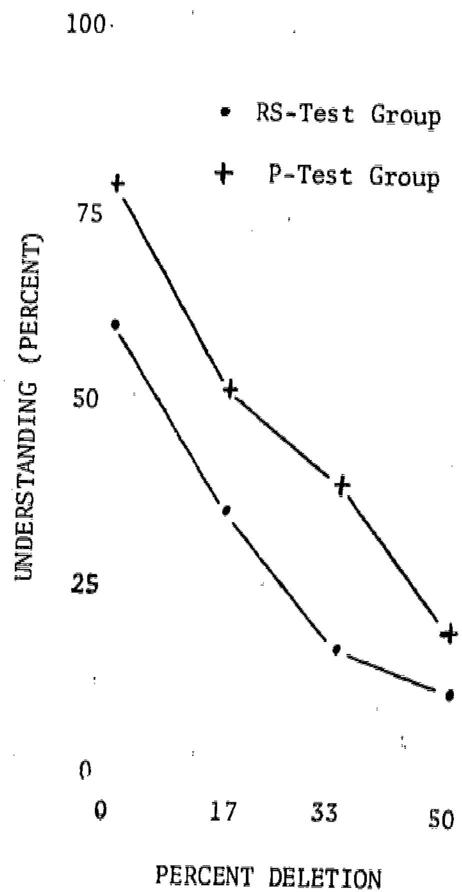


Fig. 16. Understanding as a function of the percent of passage deletion for the RS-Test Group and the P-Test Group, Experiment IID (N=40).

Even though the P-Test group reported that they understood about 15% more while they were marking less of the sentences correct, the differences between the two groups are relatively small. Thus, the data in Fig. 15 and Fig. 16 may be interpreted as indicating that the two groups were approximately equal with respect to the information they had stored. If there were differences between the two groups, they were approximately equal at each level, thus having no differential effect upon the general shapes of the RS-Test and P-Test curves.

Fig. 17 presents the RS-Test and P-Test scores as a function of the degree of deletion. The two curves are almost perfectly coincident except for the 50% deletion condition where there is a difference of 22 percentage points.

For the other four points, the average difference is only about 7 percentage points. The P-Test appears to present a more consistent decrease from the

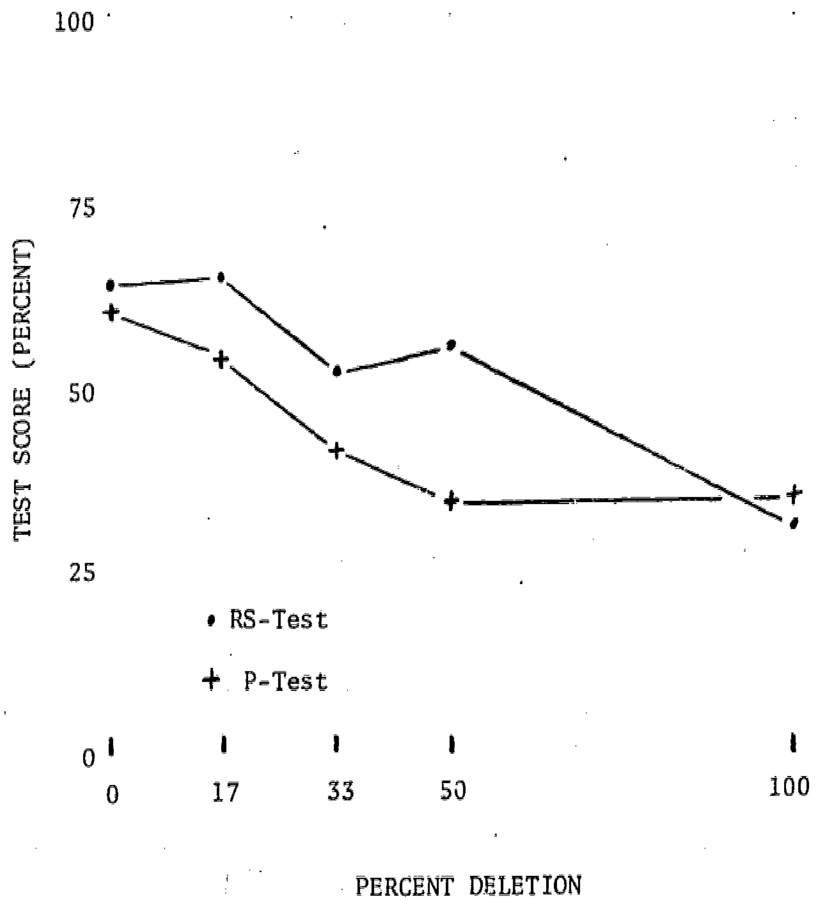


Fig. 17. Test score as a function of the percent of passage deletion for the RS-Test and the P-Test, Experiment IID (N=40).

0% deletion condition to the 50% deletion condition. However, the score on the P-Test at the 50% condition equals that of the 100% condition, and this result is inconsistent with the data in Fig. 15 and Fig. 16 which indicated that something greater than zero had been stored under the 50% deletion condition. The efficiency ratios for the reading versus the non-reading condition were as follows: RS-Test, .53; P-Test, .28. These data appear to provide no clear-cut support for the superiority of the P-Test over the RS-Test.

As in Experiment IB, the differences in Fig. 17 are probably due to randomly distributed, within-individual differences, thus making the most representative curve, the one formed by combining the data from both groups. This averaging was accomplished for both Experiment IB and Experiment IID, and these data are presented in Figure 18. The almost perfect coincidence in the two curves supports the hypothesis that all of three of the tests -- RS-Test, MC-Test; and P-Test -- are reflecting approximately the same thing. The differences

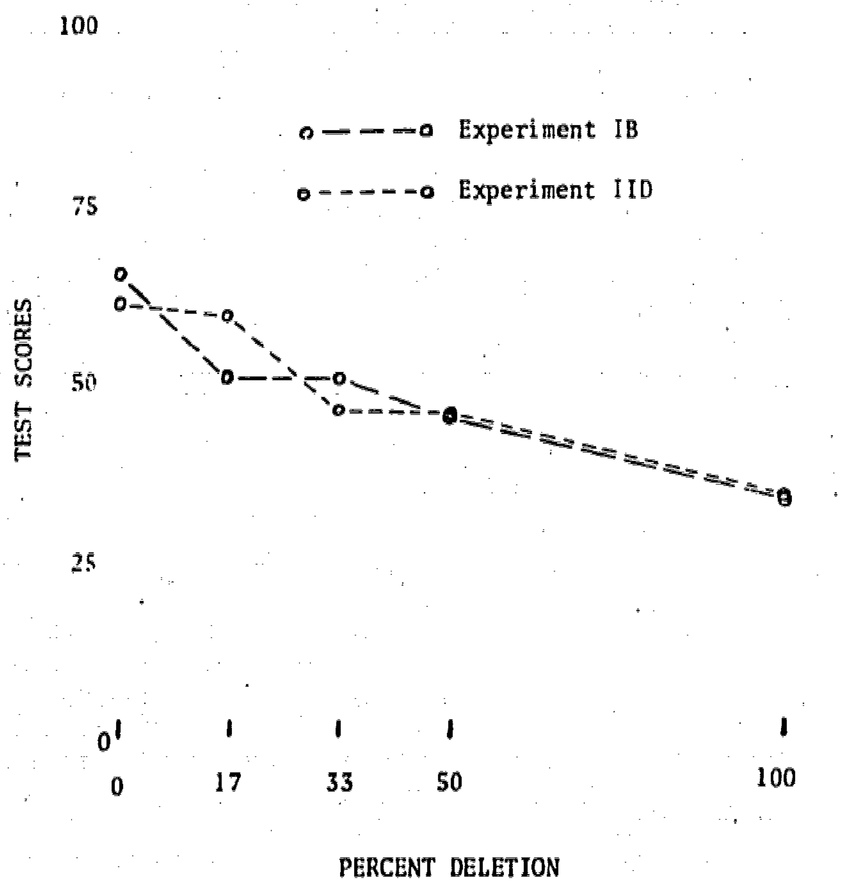


Fig. 18. Test scores as a function of the percent of passage deletion for the RS- and MC-Test data combined in Experiment IB and for the RS- and P-Test data combined in Experiment IID.

between the curves in Fig. 7 and Fig. 17 are therefore most reasonably attributed to random within individual variance that is best controlled by large sample sizes. The relatively consistent drop in scores from the 0% to the 50% deletion condition is consistent with the two understanding curves (Fig. 6 and Fig. 16) and the sentence marking curve (Fig. 15). The drop of approximately 50 percentage points in understanding is associated with about a 20% drop in the objective test scores, i.e., the efficiency ratio for all three of these tests, is about .40 in these two deletion experiments.

Conclusions. When information stored is directly manipulated by deleting words from passages, the RS-Test and the P-Test appear to be approximately equal in reflecting these changes in the primary effects of reading.

Experiment IIB.

Introduction. The primary purpose of this experiment was to investigate the effect of programmed prose upon information stored under two levels of motivation, low and high. It was hypothesized that programmed prose would facilitate the amount of information stored, i.e., amount learned, from difficult material under a low motivation condition, but be an inhibitor under a high motivation condition.

The hypothesis was tested by developing programmed prose materials on a lengthy and relatively difficult passage taken from a journal article on reading research. The regular passage and the programmed prose was then administered to college students under the two motive-incentive conditions: (a) low -- the Ss were given the prose materials and instructed to read the material carefully so that they could do their best on the test that would be administered afterwards and (b) high -- the Ss were informed that they would be paid a bonus for each question they answered correctly on the test afterwards. There was also a control group which was administered the test questions without reading. This control group was then given the regular prose, and the test was administered again afterwards. These control data were collected so as to be able to interpret the scores from the experimental groups on an absolute scale varying from what might be expected under conditions where scores on the test should be minimal and maximal.

Subjects. The Ss were the entire set of 58 college students who participated in the Set II Experiments.

Design. The overall design for the experiment is presented in Table 10. One group (A) received the regular prose under the low motivation condition. Another group (B) received the programmed prose under the low motivation condition. A third group (C) received the regular prose under the high motivation condition, and, a fourth group (D) received the programmed prose under the high motivation condition.

A fifth group (Control), not represented in Table 8, took the test under two different treatment conditions -- prior to reading the passage and after

Table 10

Design for Experiment IIB

	Motivation Condition	
	Low	High
Regular Prose	Group A	Group C
Programmed Prose	Group B	Group D

reading the passage. These two treatments were referred to as Pre-Control and Post-Control.

There was only one treatment administered per session, i.e., one group was tested each day for five consecutive days. The order of the five groups were as follows: B, A, C, D and Control.

Part of the success of the manipulation of motivation depended upon the Ss in Group A and Group B being naive with respect to the experiment. If, for example, a person in Group B, the first group, informed a friend in Group A, the second group, that he would be paid for how well he did, the person in Group A would actually belong in Group C. Since the subject pool was extremely large, it was not likely that such communication between groups would occur over a 20 hr. period. However, three factors were implemented to facilitate better control over this problem. First, the Ss in group B were informed, at the end of the session, that all groups participating in the research would recieve different bonus conditions and were asked not to divulge the nature of the experiment because it might have an adverse effect upon a friend's score. Second, the Ss in Group A were administered a questionnaire after they had been paid at the end of the experiment,inquiring as to whether they had received prior information about the bonus system. None admitted this knowledge. Third, the regular prose group, Group A, was tested second so that if there was communication between individuals in the two groups, it would benefit the regular prose group and thus tend to provide evidence against the hypothesis being tested.

Procedures and Instructions. At the outset, every group was informed that they would be given a long passage to read, and given a 100-item test on the passage afterwards. They were told that the test included 50 RS-Test items and 50 P-Test items similar to those they had taken in the preceding experiment, Experiment IA.

The Ss were told that the passage was somewhat like a test itself since there was an item every fifth word. The programmed prose goups were told to look at the first page of the passage and note that an X had already been placed in the first box. They were instructed to mark an X in every box on the remainder of the passage. The regular prose groups were told to note that all of the X's had already been marked in the correct boxes, and that they should try to read the material in a normal manner, disregarding the incorrect alternatives.

All groups were told that they would receive about 29 minutes to read the passage, and that if they finished early they should go back and read the passage again. A clock was provided indicating the amount of time remaining. All groups were also told that the scores they made on the 100-item test would be mailed to them. The high motivation groups were also told that they would receive 2¢ per each correct answer on the 100-item test.

After the initial instructions were given, the Ss were instructed to begin reading and the clock was started. The Ss were actually given 28 minutes and 42 seconds to read (an average rate of 12 programmed prose items per min.). After the time limit was up, all the groups were given the test and told that: (a) they would have 40 minutes to finish, (b) once they had finished a page and turned to the next, they could not turn back and work on a previous page, and (c) they should pace themselves by spending about 3 minutes per page and by keeping themselves informed of the time remaining.

The low motivation groups were informed at this point, i.e., after reading but prior to taking the test, that they would be paid a 2¢ bonus for each correct answer. Thus, the motive-incentive conditions were exactly the same for all groups during the administration of the 100-item test.

There was an additional difference between the programmed prose and the regular prose groups under the high motivation conditions. The programmed prose group was also informed at the outset that they would be paid 1/2¢ per each correct answer on the programmed prose materials.

For the Control Group the entire sequence of events was explained at the outset, i.e., the pre-test, the reading, and the post-test. The Control Group received regular prose, not programmed prose. This group also knew at the outset that they would receive 2¢ per each correct answer on both the pre-test and the post-test.

Passage. The passage was selected to be readable but difficult for college students. It consisted of the first 1720 words from a journal article about a computer model of reading (see Carver, 1971a). The programmed prose for the passage was developed according to the procedures given for the reading-input technique (see Carver, 1973c). There were 344 programmed prose items, i.e., one item per each five words of running text. The Fig. 1 example is an excerpt from the actual programmed prose. The passage was typed in all capital letters, i.e., simulated computer output, with 20 items per page. The items were in a vertical column justified at the right hand side of the page, i.e., there were four words and a two choice item per line making 20 lines per page. The title of the article and the three section subtitles were given, and paragraphs were signified by skipping a line. The passage was 19 pages long.

The RIDE Level difficulty of the passage was determined to be 5.3, i.e., Level 4, using the procedures described by Carver (1973d).

The passage for the regular prose condition consisted of the programmed prose passage with the correct answers to the items already marked.

Tests. One P-Test item was written for each of the first 10 sentences in the 1720 word passage. Then 10 RS-Test items were developed starting with the next sentence. After the 10 RS-Test items were completed, 10 P-Test items were written for the subsequent 10 sentences. This alternating between 10 P-Test and 10 RS-Test items continued until there were 50 items for each test. The development of these 100 items was the factor which determined the exact length of the passage.

The P-Test items were developed according to the procedures outlined in Experiment IIA and the RS-Test items were developed according to the procedures outlined in Experiment IA.

There were 10 RS-Test items per page and 5 P-Test items per page, making a total of 15 pages on the 100-item test.

Data Analysis. There are large differences between individuals with respect to their ability to store the information contained in prose materials. Also, 11 or 12 individuals per group is not nearly enough to control for these differences. Thus, the two standardized tests which were administered at the outset provided a means for control over such differences, i.e., the Basic Reading Rate Scale and the RL-4 test. The data were analyzed twice, using each standardized test as a control.

The first analysis involved the Basic Reading Rate Scale, which provides for the categorization of readers into four types -- Beginning, Good, Better, and Best. In this experiment there were individuals in each of the three higher levels -- Good, 18; Better, 36; and Best, 4. However, the distribution was not the same between the five treatment groups. For example, one group had a frequency distribution of 0, 6, 5 and another had a frequency distribution of 2, 7, 2. In order to be able to generalize the results to the most representative group of college students, the sparse data from the Ss in the Best and Good categories were deleted. Thus, one data analysis consisted of the mean scores of the Good Readers who participated in each treatment.

The second analysis involved the RL-4 test, which provided an efficiency of reading score (number right corrected for guessing, per minute). The linear regression of the test scores (i.e., RS-Test and P-Test) on the RL-4 scores was computed for each of the six treatments. Then, the average slope, $\bar{b}$, for the six treatments was found, and this value was used to adjust the mean test scores for group differences in reading ability. The above procedure was performed twice, once for the RS-Test and once for the P-Test.

Results. Both programmed prose groups were able to input the passage at a 90% or above level, i.e., the scores ranged from 90 - 99%. The median score for the low motivation, programmed prose group, Group B, was 98.5 and the corresponding value for the high motivation group, Group D, was 96.9. This high level of performance suggests that the difficulty level of the prose materials was not greater than the ability levels of all the Ss in all five groups.

Table 11 contains the means and standard deviations for each of the six treatments for each of the two types of tests, RS-Test and P-Test, and for both data analyses, BRRS and RL-4.

Table 11

Means and Standard Deviations for the Six Treatments

	Group		BRRS Analysis		RL-4 Analysis	
			Mean	S.D.	Adj. Mean	S.D.
RS-Test						
RP-Low	A	7	30.6	3.1	28.4	3.9
PP-Low	B	9	32.8	5.0	30.2	5.6
RP-High	C	7	28.7	8.3	27.9	7.5
PP-High	D	7	25.1	4.9	25.0	4.2
Pre-Cont.	E	6	24.0	4.7	22.6	5.8
Post-Cont.	E	6	31.5	7.5	28.6	7.6
P-Test						
RP-Low	A	12	29.4	5.5	26.7	6.0
PP-Low	B	12	34.6	5.7	32.9	5.8
RP-High	C	12	33.0	6.5	30.5	8.3
PP-High	D	11	26.9	5.3	25.7	4.9
Pre.Cont.	E	11	19.5	4.3	18.8	4.1
Post-Cont.	E	11	31.2	9.0	29.7	9.5

As noted earlier, The BRRS analysis means were for individuals who scored at Level 3 on the BRRS test, and the RL-4 analysis means were adjusted for group differences in the scores on the RL-4 test. The group means and standard deviations on the RL-4 test are presented in Table 12. Notice that the difference between the highest scoring group, 17.6, and the lowest scoring group, 14.6, was larger than the smallest standard deviation, 2.9, a considerable difference.

It may be noted that the scores on the RS-Test were somewhat higher than the P-Test under the Pre-Control condition. As indicated earlier, the Pre-Control and Post-Control conditions were administered so as to be able to interpret the

ERIC

Table 12

Means and Standard Deviations for the RL-4 Test

Group		N	Mean	S.D.
RP-Low	A	12	17.6	3.7
PP-Low	B	12	16.7	[illegible].8
RP-High	C	12	14.7	2.9
PP-High	D	11	16.0	4.0
Control	E	11	14.6	2.9

results on an absolute scale. In order to derive such scores, the means in Table 11 were first corrected for guessing using the following formula: rights minus one-fourth wrongs. Then, the Pre-Control means were used to calculate an absolute score using the following formula:

$$\frac{\text{Treatment Mean - Pre-Control Mean}}{\text{Post-Control Mean - Pre-Control Mean}} \times 100. \qquad (2)$$

These data for both the RS-Test and the P-Test are presented in Fig. 19 for the BRRS analysis and Fig. 20 for the RL-4 analysis.

Under the low motivation condition, the programmed prose group scored higher than the regular prose group, and under the high motivation condition the programmed prose group scored lower than the regular prose group. The RS-Test results appear to be a replication of the P-Test results and the BRRS analysis appears to be a replication of the RL-4 analysis. Thus, both analyses and both types of tests tend to present the same result even though the two tests and the two analyses are quite dissimilar.

The low motivation, programmed prose group had scores well above 100%, thus indicating that this group did better than the control group which had had two major advantages: (a) they were given an opportunity to study the questions that they would be asked in advance of their studying, and (b) they had a monetary incentive for learning.

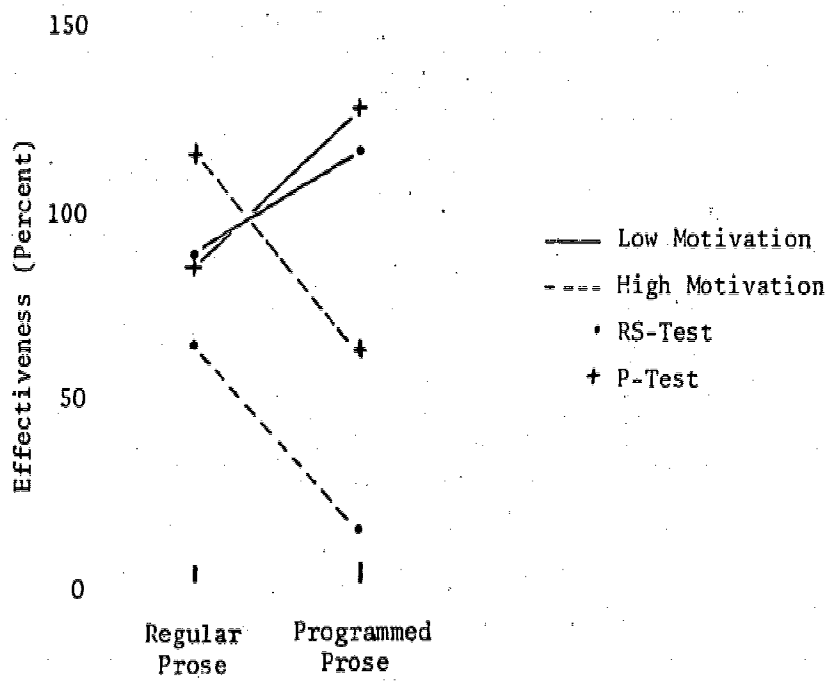


Fig. 19. Effectiveness of Regular Prose and Programmed Prose as measured by the RS-Test and the P-Test under low and high motivation conditions for the BRRS analysis

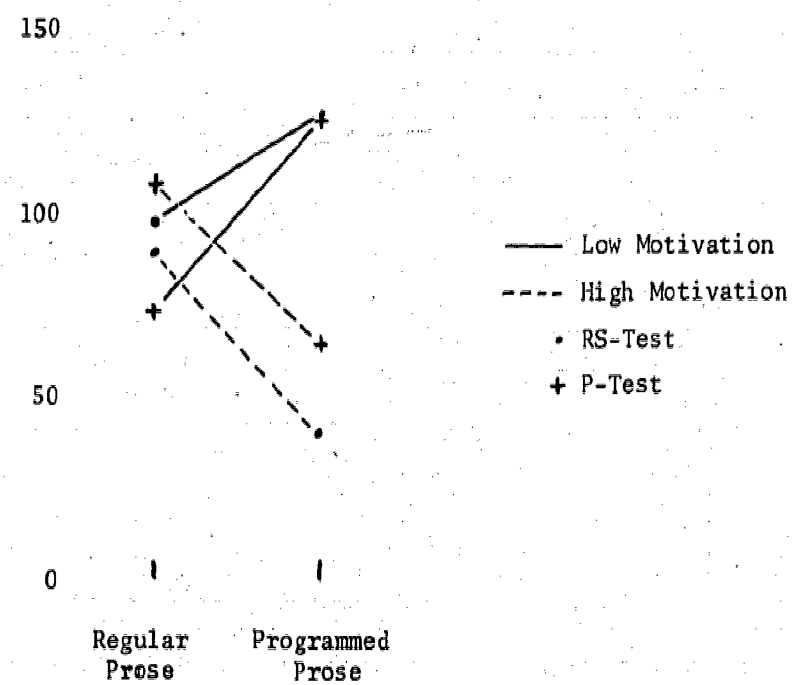


Fig. 20. Effectiveness of Regular Prose and Programmed Prose as measured by the RS-Test and the P-Test under low and high motivation conditions for the RL-4 analysis (N=58).

Discussion. These data support the hypothesis that programmed prose acts as a facilitator of attention and, as such, indirectly facilitates learning under conditions wherein attention wanes. The results were just as had been expected in this regard. When students were asked to read difficult material under conditions wherein there is very little incentive to attend to the task at hand, programmed prose was shown to increase the amount of information stored. On the other hand, when students were highly motivated to learn, i.e., they were being paid on the basis of how well they performed, programmed prose acted as an inhibitor as compared to regular prose. It appears that the use of programmed prose as a manipulator of attention and learning deserves further investigation.

Although the above results supported the interaction that was hypothesized, there are several factors that tend to qualify the result. The results held up under two drastically different analyses and two drastically different types of tests. However, the limits of this generalization need to be tested in other ways. For example, will these results generalize to lower ability individuals who are presented lower difficulty materials? Or, will these results generalize to a condition wherein many of the Ss do not have time to finish the programmed prose and go back and read much of the material again, as they did in this experiment? Furthermore, a close inspection of these data evokes a puzzling question: Why did the high motivation individuals do worse than the low motivation individuals when they were administered programmed prose? It would seem that the two high motivation groups should have done better than the two low motivation groups. Instead, the high motivation group which received regular prose, stored slightly less information than the low motivation, programmed prose group. These results could be interpreted as indicating the potency of the programmed prose, but the potency seems to interact with the motive-incentive conditions in an unexplainable manner. Rather than trying to explain these data with an anxiety factor, i.e., the high incentive groups became anxious and lost efficiency, it seems more prudent to determine whether these data can be replicated.

Finally, it should not go unnoticed that the near-perfect replication of RS-Test results and the P-Test results provides another bit of evidence for the validity of the RS-Test as a measure of the primary effects of reading. There is another result, in this regard, that supports the reliability of the RS-Test. In Experiment IC, the gain from non-reading to the forgetting plateau was 20 percentage points for a group of passages which averaged about 5.1 on the RIDE Scale. In this experiment, Experiment NB, the gain from non-reading to reading the 5.3 difficulty level passage, a roughly comparable comparison, was 19 percentage points in the BRRS analysis and 15 percentage points in the RL-4 analysis. It appears that the RS-Test will provide roughly comparable results on an absolute scale as long as the difficulty level of the passages and the ability level of the individuals are controlled.

Conclusion. These data supported the hypothesis that programmed prose is a manipulator of attention and thereby facilitates learning from prose materials under low motivation conditions,and inhibits learning under high motivation conditions. However, since those individuals who received programmed prose materials did better under low motivation conditions than they did under high motivation conditions, it seems prudent to conclude only that more should be learned about efficacy of programmed prose and its interaction with motivation.

Summary and Conclusions

Phase I

The purpose of Phase I was to investigate the validity of the reading-storage test as a measure of the primary effect of reading, whether the primary effect is called learning, information stored, understanding, or comprehension. Six different experiments were conducted which were directly relevant to this purpose.

In the first three experiments, the reading-storage test was compared to a modified version of the cloze test. Validity was investigated in one experiment by investigating the sensitivity of the tests to reading prose, as compared to not reading, at four levels of difficulty. In another experiment, the effect of reading was manipulated by deleting various percents of the words from prose, and the tests were compared with respect to the measurement of these changes. In the third experiment, the sensitivity of the tests to forgetting was studied. The reading-storage type of test appeared to be just as valid if not more valid than the modified cloze type of test as a measure of the primary effects of reading. The cloze procedure represents the only other technique besides the reading-storage procedure which is completely objective from a test development standpoint. Also, none of the cloze techniques appear to be more valid than the reading-storage technique, while the reading-storage technique appears to be about twice as valid as the regular cloze technique. Since the reading-storage test appears to be just as valid as the most valid version of the cloze test, since the reading-storage test can be scored objectively while the cloze test requires subjective scoring, and since the cloze test appears to be the only other type of objectively developed test appropriate for prose materials, it has been concluded that the reading-storage technique represents a better dependent variable than any other existing objective technique for measuring the primary effects of reading.

In the Set II Experiments, the validity of the reading-storage test was investigated further by comparing it to a technique which has more intuitive appeal, the paraphrase technique. In one experiment, comprehension was manipulated by using special paragraphs which were highly difficult to comprehend without certain context cues. It was hypothesized that if the reading-storage test was primarily sensitive to the memorization of words instead of the primary effects of reading, then the score on the paragraphs given without context cues should be the same as scores on the test given with context cues. Also, if memorization is what the reading-storage test primarily measures, then the reading-storage scores should be much higher on the test given without context cues than they are on a test that is given without the opportunity to read the paragraph first. The results indicated that there was slightly more gain, on the average, due to the assumed comprehension conditions than there was due to the assumed memorization conditions. More importantly, however, the reading-storage test was more sensitive to the assumed comprehension effects than was the paraphrase test, and it was less sensitive to the assumed memorization effects than was the paraphrase test.

Two other experiments in Set II were replications of the two previously described experiments,where validity was explored by comparing reading scores with non-reading scores and by manipulating the effects of reading by deleting varying percents of the words.

The paraphrase test was found to be somewhat more valid than the reading-storage test in one experiment but the reading-storage test was found to provide much more consistent or reliable results than the paraphrase test. In the other experiment, the two tests appeared to be equally effective.

Considering the results from all three experiments, the reading-storage test appears to be just as valid a dependent variable a is the intuitively more appealing, paraphrase test. Since the paraphrase test is developed subjectively, it is always questionable to generalize the results beyond the experimenter who developed the test. Therefore, if the paraphrase technique is not shown to be consistently more valid than the reading-storage technique, it appears reasonable to conclude that the reading-storage technique is more valid as a general method for measuring the primary effects of reading. The completely objective, reading-storage test appears to provide a better technique than its two closest competitors, i.e., the cloze technique which is developed objectively but scored subjectively, and the paraphrase technique which is developed subjectively, but may be scored objectively.

The Phase II data, which was collected primarily to investigate the effectiveness of programmed prose, also provides support for the validity of the reading-storage test because the paraphrase test results and the reading-storage test results were essentially equivalent.

Phase II.

The primary purpose of Phase II was to investigate the effect of programmed prose upon learning. It was hypothesized that programmed prose would act to increase attention in low motivation conditions, where attention would be expected to wane, and thereby facilitate learning. It was also hypothesized that programmed prose would act as a distractor in high motivation conditions, where attention would not be expected to wane, and thereby inhibit learning. These hypotheses were tested by administering regular prose and programmed prose under low and high motivation conditions. Under the low motivation conditions, the Ss were simply asked to do their best. Under the high motivation conditions, the Ss were told that they would be paid on the basis of how well they did. The results supported the hypotheses, i.e., programmed prose facilitated learning under the low motivation conditions and inhibited learning under the high motivation conditions. However, the scores from the low motivation, programmed prose group were unexpectedly higher than the scores from the high motivation, programmed prose group. It was concluded that the efficacy of programmed prose and its interaction with motivation deserves to be investigated further using different techniques for manipulating motivation, and using different levels of prose difficulty and individual ability.

References

Anderson, Richard C. Control of student mediating processes during verbal learning and instruction. Review of Educational Research, 1970, 40 349-370.

Anderson, Richard C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42, 145-170.

Bormuth, John R. Factor validity of cloze tests as measures of reading comprehension ability. Reading Research Quarterly, 1969, 4, 358-365.(a)

Bormuth, John R. Development of readability analyses. U.S.O.E. Final Report, Proj. No. 7-0052, Contract No. OEC-3-7-070052-0326, University of Chicago, March, 1969.(b)

Bradley, J. V. Complete counter balancing of immediate sequential effects in a Latin square design. American Statistical Association Journal, 1958, 53,525-528.

Bransford, J. D. & Johnson, M. K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 1972, Dec.

Carver, Ronald P. Reading Progress Scale. Silver Spring, Md. : Revrac Publications, 1970.

Carver, Ronald P. A computer model of reading and its implications for measurement and research. Reading Research Quarterly, 1971, 6, 449-471.(a)

Carver, Ronald P. Procedures for constructing a variety of information-processing measures appropriate for prose materials. Silver Spring, Md. :Revrac Publications, 1971.(b)

Carver, Ronald P. A critical review of "mathemagenic" behaviors and the effect of questions upon the retention of prose materials. Journal of Reading Behavior, 1972, 4, 93-119.(a)

Carver, Ronald P. Evidence for the invalidity of the Miller-Coleman readability scale. Journal of Reading Behavior, 1972, 4, 42-47.(b)

Carver, Ronald P. Understanding, information-processing, and learning from prose materials. Journal of Educational Psychology, 1973, in press.(a)

Carver, Ronald P. Toward a comprehensive theory of reading and prose rauding. Unpublished manuscript, 1973.(b)

Carver, Ronald P. Revised procedures for developing reading-input and reading storage tests. Unpublished manuscript, 1973.(c)

Carver, Ronald P. A measure of prose difficulty: the RIDE Scale. Unpublished manuscript, 1973.(d)

Coleman, Edmund B. & Miller, Gerald R. A measure of information gained during prose learning. Reading Research Quarterly, 1968, 369-386.

Danks, Joseph H. Grammaticalness and meaningfulness in the comprehension of sentences. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 687-696.

Frijda, N. H. Simulation of long-term memory. Psychological Bulletin, 1972, 77, 1-31.

Jongsma, Eugene. The cloze procedure as a teaching technique. Newark, Delaware: ERIC/CRIER and the International Reading Association, 1971.

King, D.J. Influence of interitem interval in the learning of connected discourse. Journal of Experimental Psychology, 1971, 87, 132-134.

Miller, Gerald R. & Coleman, Edmund B. The measurement of reading speed and the obligation to generalize to a population of reading materials. Journal of Reading Behavior, 1972, 4, 43-48.

Newell, A. & Simon, H. A. Overview: Memory and process in concept formation. In B. Kleinmuntz (Ed.) Concepts and the structure of memory. New York: Wiley, 1967, Pp. 241-262.

Pribram, Karl H. Meaning in education. Invited address at the annual meeting of the American Educational Research Association, Chicago, 1968.

Robinson, H. M. Sequential development of reading abilities. Supplementary Education Monographs #90. Chicago: University of Chicago Press, 1960.

Rothkopf, Ernst Z. Some theoretical and experimental approaches to problems in written instruction. In J. D. Krumboltz (Ed.) Learning and the instruction process, Chicago: Rand McNally, 1965. [Pp. 193-221]

Rothkopf, Ernst Z. The concept of mathemagenic activities. Review of Educational Research, 1970, 40, 325-336.

Schwartz, D. Sparkman, J. P. & Deese, J. The process of understanding and judgments of comprehensibility. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 87-93.

Skinner, Burris F. Verbal behavior. New York: Appleton-Century-Crofts, 1957.

Spinelli, D. N. OCCAM: A computer model for a content addressable memory in the central nervous system. In K. H. Pribram & D. E. Broadbent (Eds.) Biology of Memory. New York: Academic Press, 1970. [Pp. 293-306]

Taylor, Wilson L. Cloze procedure: A new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

Thorndike, Edward L. Reading as reasoning: A study of mistakes in paragraph reading. Journal of Educational Psychology, 1917, 8, 323-332.

DISTRIBUTION LIST

NAVY

4 Dr. Marshall J. Farr
Director, Personnel and Training Research Programs
Office of Naval Research
Arlington, Virginia 22217

1 Director
ONR Branch Office
495 Summer Street
Boston, Massachusetts 02210

1 Director
ONR Branch Office
1030 East Green Street
Pasadena, California 91101

1 Director
ONR Branch Office
536 South Clark Street
Chicago, Illinois 60605

1 Office of Naval Research
Area Office
207 West 24th Street
New York, New York 10011

1 Office of Naval Research
Area Office
1076 Mission Street
San Francisco, California 94103

1 Commander
Operational Test and Evaluation Force
U. S. Naval Base
Norfolk, Virginia 23511

6 Director
Naval Research Laboratory
Code 2627
Washington, D.C. 20390

12 Defense Documentation Center
Cameron Station, Building 5
5010 Duke Street
Alexandria, Virginia 22314

1 Chairman
Behavioral Science Department
Naval Command and Management Division
U. S. Naval Academy
Luce Hall
Annapolis, Maryland 21402

1 Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, Tennessee 38054
ATTN: Dr. G. D. Mayo

1 Chief of Naval Training
Naval Air Station
Pensacola, Florida 32508
ATTN: CAPT Allen E. McMichael

1 Chief
Bureau of Medicine and Surgery
Code 513
Washington, D.C. 20390

1 Chief
Bureau of Medicine and Surgery
Research Division (Code 713)
Department of the Navy
Washington, D.C. 20390

1 Commander Naval Air Research
Naval Air Station
Glenview, Illinois 60026

1 Commander
Naval Air Systems Command
Navy Department, AIR-413C
Washington, D.C. 20360

1 Commander
Submarine Development Group Two
Fleet Post Office
New York, New York 09501

1 Commanding Officer
Naval Air Technical Training Center
Jacksonville, Florida 32213

Distribution List (Cont'd)

1 Commanding Officer
Naval Personnel and Training
Research Laboratory
San Diego, California 92152

1 Commanding Officer
Service School Command
U. S. Naval Training Center
San Diego, California 92133
ATTN: Code 303

1 Head, Personnel Measurement Staff
Capital Area Personnel Service Office
Ballston Tower #2, Room 1204
801 N. Randolph Street
Arlington, Virginia 22203

1 Program Coordinator
Bureau of Medicine and Surgery
(Code 71G)
Department of the Navy
Washington, D.C. 20390

1 Research Director, Code 06
Research and Evaluation Department
U. S. Naval Examining Center
Building 2711 - Green Bay Area
Great Lakes, Illinois 60088
ATTN: C. S. Winiewicz

1 Superintendent
Naval Postgraduate School
Monterey, California 93940
ATTN: Library (Code 2124)

1 Technical Director
Naval Personnel Research and
Development Laboratory
Washington Navy Yard
Building 200
Washington, D. C. 20390

1 Technical Director
Personnel Research Division
Bureau of Naval Personnel
Washington, D. C. 20370

1 Technical Library (Pers-11B)
Bureau of Naval Personnel
Department of the Navy
Washington, D. C. 20360

1 Technical Library
Naval Ship Systems Command
National Center
Building 3 Room 3
S-08
Washington, D. C. 20360

1 Technical Reference Library
Naval Medical Research Institute
National Naval Medical Center
Bethesda, Maryland 20014

1 Behavioral Sciences Department
Naval Medical Research Institute
National Naval Medical Center
Bethesda, Maryland 20014

1 COL George Caridakis
Director, Office of Manpower Utilization
Headquarters, Marine Corps (A01H)
MCB
Quantico, Virginia 22134

1 Special Assistant for Research
and Studies
OASN (M&RA)
The Pentagon, Room 4E794
Washington, D. C. 20350

1 Mr. George N. Graine
Naval Ship Systems Command
(SHIPS 03H)
Department of the Navy
Washington, D. C. 20360

1 CDR Richard L. Martin, USN
COMFAIRMIRAMAR F-14
NAS Miramar, California 92145

1 Mr. Lee Miller (AIR 413E)
Naval Air Systems Command
5600 Columbia Pike
Falls Church, Virginia 22042

Distribution List (Cont'd)

1 Dr. James J. Regan
Code 55
Naval Training Device Center
Orlando, Florida 32813

1 Dr. A. L. Slafkosky
Scientific Advisor (Code Ax)
Commandant of the Marine Corps
Washington, D. C. 20380

1 LCDR Charles J. Theisen, Jr., MSC, USN
CSOT
Naval Air Development Center
Warminster, Pennsylvania 18974

ARMY

1 Behavioral Sciences Division
Office of Chief of Research and Development
Department of the Army
Washington, D. C. 20310

1 U.S. Army Behavior and Systems Research Laboratory
Rosslyn Commonwealth Building, Room 239
1300 Wilson Boulevard
Arlington, Virginia 22209

1 Director of Research
U.S. Army Armor Human Research Unit
ATTN: Library
Building 2422 Morade Street
Fort Knox, Kentucky 40121

1 COMMANDANT
U.S. Army Adjutant General School
Fort Benjamin Harrison, Indiana 46216
ATTN: ATSAG-EA

1 Commanding Officer
ATTN: LTC Montgomery
USACDC - PASA
Ft. Benjamin Harrison, Indiana 46249

1 Director
Behavioral Sciences Laboratory
U.S. Army Research Institute of Environmental Medicine
Natick, Massachusetts 01760

1 Commandant
United States Army Infantry School
ATTN: ATSIN-H
Fort Benning, Georgia 31905

1 U.S. Army Research Institute
Room 239
Commonwealth Building
1300 Wilson Boulevard
Arlington, Virginia 22209
ATTN: Dr. R. Dusek

1 Mr. Edmund Fuchs
BESRL
Commonwealth Building, Room 239
1320 Wilson Boulevard
Arlington, Virginia 22209

AIR FORCE

1 AFHRL (TR/Dr. G. A. Eckstrand)
Wright-Patterson Air Force Base
Ohio 45433

1 AFHRL (TRT/Dr. Ross L. Morgan)
Wright-Patterson Air Force Base
Ohio 45433

1 AFHRL/MD
701 Prince Street
Room 200
Alexandria, Virginia 22314

1 AFOSR (NL)
1400 Wilson Boulevard
Arlington, Virginia 22209

1 COMMANDANT
USAF School of Aerospace Medicine
ATTN: Aeromedical Library (SCL-4)
Brooks AFB, Texas 78235

Distribution List (Cont'd)

1 Personnel Research Division
AFHRL
Lackland Air Force Base
San Antonio, Texas 78236

1 Headquarters, U.S. Air Force
Chief, Personnel Research and Analysis Division (AF/DPXY)
Washington, D. C. 20330

1 Research and Analysis Division
AF/DPXYR Room 4C200
Washington, D. C. 20330

1 Headquarters Electronic Systems Division
ATTN: Dr. Sylvia R. Mayer/MCIT
LG Hanscom Field
Bedford, Massachusetts 01730

1 CAPT Jack Thorpe USAF
Department of Psychology
Bowling Green State University
Bowling Green, Ohio 43403

DOD

1 Mr. William J. Stormer
DOD Computer Institute
Washington Navy Yard
Building 175
Washington, D. C. 20390

1 Mr. Joseph J. Cowan, Chief
Psychological Research Branch (P-1)
U.S. Coast Guard Headquarters
400 Seventh Street, SW
Washington, D. C. 20590

1 Dr. Charles Ullman
Chief of Counseling, Training Programs
OSD (M&RA)
The Pentagon, Room 2C252
Washington, D. C. 20301

OTHER GOVERNMENT

1 Dr. Alvin E. Goins, Chief
Personality and Cognition Research Section
Behavioral Sciences Research Branch
National Institute of Mental Health
5600 Fishers Lane
Rockville, Maryland 20852

1 Dr. Andrew R. Molnar
Computer Innovation in Education Section
Office of Computing Activities
National Science Foundation
Washington, D. C. 20550

1 Office of Computer Information
Center for Computer Sciences and Technology
National Bureau of Standards
Washington, D. C. 20234

MISCELLANEOUS

1 Dr. Scarvia Anderson
Executive Director for Special Development
Educational Testing Service
Princeton, New Jersey 08540

1 Professor John Annett
The Open University
Waltonteale, BLETCHLEY
Bucks, ENGLAND

1 Dr. Richard C. Atkinson
Department of Psychology
Stanford University
Stanford, California 94305

1 Dr. Bernard M. Bass
University of Rochester
Management Research Center
Rochester, New York 14627

Distribution List (Cont'd)

1 Professor Mats Bjorkman
University of Umea
Department of Psychology
Radhuseplanaden 2
S-902 47 UMEA/SWEDEN

1 Mr. H. Dean Brown
Stanford Research Institute
333 Ravenswood Avenue
Menlo Park, California 94025

1 Dr. Jaime Carbonell
Bolt Beranek and Newman
50 Moulton Street
Cambridge, Massachusetts 02138

1 Dr. Kenneth E. Clark
University of Rochester
College of Arts and Sciences
River Campus Station
Rochester, New York 14627

1 ERIC
Processing and Reference Facility
4833 Rugby Avenue
Bethesda, Maryland 20014

1 Dr. Victor Fields
Department of Psychology
Montgomery College
Rockville, Maryland 20850

1 Dr. Robert Glaser
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, Pennsylvania 15213

1 Dr. Albert S. Glickman
American Institutes for Research
8555 Sixteenth Street
Silver Spring, Maryland 20910

1 Dr. Bert Green
Department of Psychology
Johns Hopkins University
Baltimore, Maryland 21218

1 Dr. Duncan N. Hansen
Center for Computer-Assisted Instruction
Florida State University
Tallahassee, Florida 32306

1 Dr. M. D. Havron
Human Sciences Research, Inc.
Westgate Industrial Park
7710 Old Springhouse Road
McLean, Virginia 22101

1 Human Resources Research Organization
Division #3
Post Office Box 5787
Presidio of Monterey, California 93940

1 Human Resources Research Organization
Division #4, Infantry
Post Office Box 2086
Fort Benning, Georgia 31905

1 Human Resources Research Organization
Division #5, Air Defense
Post Office Box 6057
Fort Bliss, Texas 79916

1 Library
HumRRO Division Number 6
P.O. Box 428
Fort Rucker, Alabama 36360

1 Dr. Lawrence B. Johnson
Lawrence Johnson and Associates, Inc.
2001 "S" Street, NW
Suite 502
Washington, D. C. 20009

1 Dr. Norman J. Johnson
Associate Professor of Social Policy
School of Urban and Public Affairs
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

1 Dr. Roger A. Kaufman
Graduate School of Human Behavi
U.S. International University
8655 E. Pomerada Road

Distribution List (Cont'd)

1 Dr. E. J. McCormick
Department of Psychological Sciences
Purdue University
Lafayette, Indiana 47907

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
Santa Barbara Research Park
6780 Cortona Drive
Goleta, California 93017

1 Mr. Luigi Petrullo
2431 North Edgewood Street
Arlington, Virginia 22207

1 Dr. Robert D. Pritchard
Assistant Professor of Psychology
Purdue University
Lafayette, Indiana 47907

1 Dr. Diana M. Ramsey-Klee
R-K Research & System Design
3947 Ridgemont Drive
Malibu, California 90265

1 Dr. Joseph W. Rigney
Behavioral Technology Laboratories
University of Southern California
3717 South Grand
Los Angeles, California 90007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, Maryland 20850

1 Dr. George E. Rowland
Rowland and Company, Inc.
Post Office Box 61
Haddonfield, New Jersey 08033

1 Dr. Benjamin Schneider
Department of Psychology
University of Maryland
College Park, Maryland 20742

1 Dr. Robert J. Seidel
Human Resources Research Organization
300 N. Washington Street
Alexandria, Virginia 22314

1 Dr. Arthur I. Siegel
Applied Psychological Services
Science Center
404 East Lancaster Avenue
Wayne, Pennsylvania 19087

1 Dr. Benton J. Underwood
Department of Psychology
Northwestern University
Evanston, Illinois 60201

1 Dr. David Weiss
University of Minnesota
Department of Psychology
Elliott Hall
Minneapolis, Minnesota 55455

1 Mr. Edmund C. Berkeley
Berkeley Enterprises, Inc.
815 Washington Street
Newtonville, Massachusetts 02160

1 Mr. Dennis J. Sullivan
Training Department
Hughes Aircraft Co.
P.O. Box 90515
Los Angeles, California 90009

1 Dr. Frank J. Harris
Chief, Training Division
U.S. Army Research Institute
1300 Wilson Boulevard
Arlington, Virginia 22209

1 LCOL Austin W. Kibler, Director
Human Resources Research Office
ARPA
1400 Wilson Boulevard
Arlington, Virginia 22209

Distribution List (Cont'd)

1 Dr. Anita S. West
Denver Research Institute
University of Denver
Denver, Colorado 80210

1 Century Research Corporation
4113 Lee Highway
Arlington, Virginia 22207

1 Dr. Dan Landis, Director
Urban Education and Cross-Culture
Skills
University City
Science Center
3508 Science Center
Philadelphia, Pennsylvania 19104

1 Chief of Naval Training Support
(Code N-21)
Building 45
Naval Air Station
Pensacola, Florida 32508

AMERICAN INSTITUTES FOR RESEARCH

Corporate Officers

John C. Flanagan, PhD
President

Brent Baxter, PhD
Executive Vice President—
Development

Edwin A. Fleishman, PhD
Senior Vice President

Paul A. Schwarz, PhD
Executive Vice President—
Operations

Board of Directors

John C. Flanagan, PhD, Chairman

Frederick B. Davis, EdD

Robert M. Gagné, PhD

Paul Horst, PhD

James M. Houston, LLB

S. Rains Wallace, PhD

Research Offices

ASIA/PACIFIC
Bangkok, Thailand
APO San Francisco 96346

KENSINGTON
10605 Concord Street
Kensington, Maryland 20795

PALO ALTO
P.O. Box 1113
Palo Alto, California 94302

PITTSBURGH
710 Chatham Center Office Building
Pittsburgh, Pennsylvania 15219

WASHINGTON
8555 Sixteenth Street
Silver Spring, Maryland 20910

Corporate Offices

PALO ALTO
P.O. Box 1113
Palo Alto, California 94302

PITTSBURGH
710 Chatham Center Office Building
Pittsburgh, Pennsylvania 15219

WASHINGTON
1775 K Street, N.W.
Washington, D.C. 20006