

Negative microfilm
received from university

65-5308

GREENE, Frank Pierrepont, 1933-
A MODIFIED CLOZE PROCEDURE FOR
ASSESSING ADULT READING COMPREHENSION.

The University of Michigan, Ph.D., 1964
Education, psychology

University Microfilms, Inc., Ann Arbor, Michigan

A MODIFIED CLOZE PROCEDURE FOR ASSESSING ADULT
READING COMPREHENSION

Frank Pierrepont Greene

A Dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the
University of Michigan
1964

Doctoral Committee:

Associate Professor Donald E. P. Smith, Chairman
Assistant Professor Daniel H. Carson
Associate Professor James W. Downer
Associate Professor Melvin Manis
Associate Professor Frank B. Womer

Acknowledgements

The inexhaustible supply of both comfort and constructive criticism emanating from my doctoral committee is gratefully acknowledged.

Dr. Donald E. P. Smith has provided succor extending vastly beyond this study.

My wife, Carol, has been a source of great strength and joy. Our children have provided much incentive and sympathetic understanding of Daddy's work.

Table of Contents

Acknowledgements	Page 11
List of Tables	v
Chapter	
I. The Cloze Procedure as a Technique for Assessing Reading Comprehension	1
A. Description of Cloze Procedures	1
B. Assessing Reading Comprehension	3
C. Studies of the Cloze Procedure	4
1. Genesis: Taylor	5
2. Augmented Recall: Jenkinson	8
3. Reading Comprehension: Rankin	10
4. Verbal Abilities: Fletcher	15
5. Aural Closure: Weaver	19
6. Readability: Bormuth	21
7. Language Patterns: Ruddell	24
D. Measurement Considerations	27
1. Prior Closure Tests	27
2. Comprehension as Process or Product	28
3. Factor Analyses of Comprehension	30
4. Measuring Vocabulary	31
E. Problems in Using the Cloze Procedure	32
1. Length of Cloze Tests	32
2. Scoring Methods	33
3. Deletion Restrictions	35
4. Problems of Direct Observation	37
II. Evaluation of Alternative Cloze Procedures for Assessing Adult Reading Comprehension	40
A. Development of Hypotheses	40
B. Setting and Subjects	47
C. Materials Used	48

Table of Contents (continued)

Chapter	Page
III. Validity Analysis of Cloze Test Procedures	58
IV. Evaluation of the Cloze Procedure Modification Results	72
Some Unresolved Problems	76
Appendices	
Appendix A - The <u>Michigan Adult Reading Test</u>	79
Appendix B - Full data analysis tables	119
Bibliography	123

List of Tables

Table	Page
1. Relationship between Scores on Cloze Tests of Two Deletion Patterns and Scores on Sub-tests of the <u>Diagnostic Reading Test</u> (after Rankin, 1958)	11
2. Relationships between Pre-reading and Post-reading Cloze Test Scores (Noun-Verb Deletions) and Scores of the <u>Diagnostic Reading Test</u> (after Rankin, 1958)	12
3. Distribution of Item Discrimination Indices (r) (after Fletcher, 1959)	17
4. Distribution of Item Difficulty Indices (p) (after Fletcher, 1959)	18
5. Distribution of Item Discrimination Indices (r) (after Bormuth, 1962)	23
6. Distribution of Item Difficulty Indices (p) (after Bormuth, 1962)	24
7. Reliabilities (Split-half) of Cloze Tests (after Ruddell, 1963)	25
8. Validity Coefficients of Cloze Tests and the <u>Stanford Achievement Test</u> , Paragraph Section (after Ruddell, 1963)	26
9. Correlation Coefficients of the <u>Michigan Adult Reading Test</u> and Other Reading Tests	58
10. Comparison of Two Forms (B and C) of the <u>Michigan Adult Reading Test</u> Constructed by the Standard Deletion Procedure	60
11. Reliability Coefficients of Standard and Modified Cloze Tests	61
12. Item Discrimination Indices (r) of Two Kinds of Cloze Tests (Percentages of Items)	62
13. Item Difficulty Indices (p) of Cloze Tests (Percentages of Items)	64

List of Tables (continued)

Table	Page
14. Cloze Performances with and without Reading Original Passage Prior to Taking Tests	66
15. Relationship of Reading Rate to Cloze Scores	68
16. Relationships of Two Kinds of Vocabulary Tests and Two Cloze Tests	70

Chapter I

The Cloze Procedure as a Technique for Assessing Reading Comprehension

This is a study of adult reading comprehension as measured by the cloze procedure. Two problems were investigated. The first problem was an evaluation of certain modifications in the cloze procedure. It was an attempt to increase the validity of this method for assessing reading comprehension. The second problem was that of identifying some of the component sub-skills of reading comprehension as evidenced by successful cloze performances.

A. Description of Cloze Procedures

The standard cloze (stdC) procedure consists of deleting every nth word from a reading passage. The deleted words are replaced by blanks of some constant length. The reader is asked to fill in the words missing from the passage. His "reading comprehension" is inferred from his relative success in reconstituting the passage from the remaining parts. Under the standard procedure it is permissible to specify the class or classes of words eligible for deletion and the percentage of words to be deleted. The key feature in the standard procedure is the automatic way in which the deletions are made. Once the eligible words and deletion rate have been set, the deletion process must be automatic - every nth word removed.

In the present study, no restrictions were placed on eligible word classes for the standard cloze procedure. Thus, with any-word eligible, a die was rolled to choose the first deletion and then every 12th word following was left out.

On the other hand, the modified cloze (modC) procedure used in the present study differed in two respects from the standard. First, the classes of words eligible for deletion were limited to nouns, verbs, adverbs, and adjectives (content words) as opposed to the any-word eligibility of the stdC. This limitation to content word deletions needs to be recognized, but, as noted above, such class restrictions are acceptable within the realm of the stdC. The second way in which the modC differed from the stdC concerns the deletion pattern. This is the critical difference in that the modC procedure followed a rational, non-random deletion pattern. The modC averaged one word in twelve, but each potential word was evaluated for its reproducibility rather than being deleted automatically. That is, each word deleted was considered to have sufficient lexical and/or syntactic redundancy that positive identification of the missing word could be made by a superior reader.

For both cloze procedures, only single words were deleted. A word is defined by the space at either end. Thus, dates, numbers, and hyphenated words are all single words. A cloze response is what the reader writes in the space provided.

Only exact replacements are scored as correct. The cloze test is the collection of all the cloze items, with the test score being expressed as the percentage of the total possible items which are correctly filled.

B. Assessing Reading Comprehension

The act or process of comprehending is not open to direct measurement. As a result, reading comprehension is defined as that which is measured by a reading test. In evaluating tests of reading comprehension, therefore, we need to examine the adequacy of the test as an operational definition of the inferred comprehension behavior, in addition to the usual questions of test validity and reliability.

These two aspects of the study of comprehension formed the two stages of this study. The question of test efficiency (concurrent validity, reliability and item homogeneity) forms the first step. The adequacy of cloze test results as an operational definition of reading comprehension form the second step. It was evaluated by summing certain sub-skills to approximate the cloze test results.

At both stages, differences resulting from modification of the standard cloze procedure are evaluated with the aim of increasing the efficiency of this measure of adult reading comprehension.

C. Studies of the Cloze Procedure

In 1953, Taylor published an article on the use of the cloze technique for assessing readability. This was the first use of "cloze" to refer to the deletion process being considered here. Since then Taylor (1954), Jenkinson (1957), Rankin (1958), Fletcher (1959), Weaver (1961), Bormuth (1962) and Ruddell (1963) have all produced doctoral research using the cloze procedure as the major assessment technique. Summaries of these theses will include selected methods and data. These summaries will then be used as a background against which to discuss problems of relevance to this study. (Referencing and general style follows the American Psychological Association Publication Manual, 1957 Revision.)

The several names for the grammatical classes used in the following studies can be equated. "Content," "lexical" and "form-class" all refer to nouns, verbs, adverbs, and sometimes also to adjectives. They carry the semantic content of the passage. "Function," "structural," and "connective" subsume all other possible grammatical classes. They concern the relationships between the words in a passage; that is, they include the syntactical signals of the passage. Whatever labels were used in the various studies, they will be referred to herein as "content" and "function" words. An any-word deletion procedure would delete words of both classes.

1. Genesis: Taylor.

Application of Cloze and Entropy Measures to the Study of Contextual Constraint in Samples of Continuous Prose (1954).

Taylor discussed his development of the cloze procedure and his initial experimental findings. He used college freshmen who took four 35 item cloze tests in 40-45 minutes. Two of the pages combined to give a 70 item easy test and two combined to give a 70 item hard test. The assignment of difficulty was made on the basis of the articles from which the tests were made and not on the basis of how the cloze tests were constructed. Papers that had many unattempted blanks were excluded from analysis.

In his thesis, as in his other papers, Taylor was concerned with the materials and process rather than with the individual. Taylor uses readability to refer to differences between passages when responded to by a modal subject. Comprehension, as viewed in the present study, refers to differences between individuals with the material held constant. The same phenomena are involved in readability studies as are those in comprehension studies. Readability may be viewed as the average comprehension score for some group.

Taylor used an any word deletion pattern of every 5th word. He argued that to restrict the class of eligible words is to ignore real differences of word class usage between passages which should affect readability. He was not concerned

with the individual ratings. He was interested in the group's ability to "read" the passage; therefore, only exact replacements were scored as correct.

His study resulted in the following findings of interest to the present study.

Subjects who had previously studied either of the original passages from which the cloze passages were drawn made significantly higher cloze scores on that passage than did subjects not so exposed. The studying of the original sources had been done in regular English classes some time earlier in the semester rather than just prior to the cloze testing, so that Taylor concluded that "delayed" memory does affect cloze scores.

While the cloze scores of the easier passage were significantly higher than the scores on the more difficult one, the order of administration (easy-hard or hard-easy) did not make a difference in the cloze scores.

A high negative relationship ($r = -.87$) existed between the cloze score of each item and the uncertainty score for that same item. The uncertainty measure has to do with the extent of agreement on the deleted word, while the cloze score has to do with ability to agree with the author's choice of the deleted word. Note that these are not the same measures. An item may have a low cloze score and yet have an uncertainty

score from low to high. If all the subjects agree on the word to be filled in, the uncertainty is low, even though that word may not match the author's choice, in which case the cloze score will also be low. If the subjects do not agree on the missing word, not only will the uncertainty be high, but the cloze score must be low, since there was a wide choice of words only one of which is cloze-correct.

Neither validities nor reliabilities were reported since Taylor's interest was on intra-procedure differences. He used the results of the group who took the easy test as the first task for detailed item analysis ($N = 50$). The average cloze score for the easy essay was 54.6% on 70 items. For the hard passage the cloze score had averaged 37.0%. Although the distribution of the cloze scores was approximately normal, the distribution of item difficulties was abnormal, being essentially rectangular. The distribution of the item difficulties for the easy and hard tests looked very close, except that the easy test had a few items pile up at the low difficulty end, while the hard test had a few items pile up at the high difficulty end.

Lastly, Taylor looked at the number of items necessary to produce a stable standard cloze score. Random factors may make the particular set of words chosen different from another set that might have been chosen. That is, a different

form derived by starting with a different word might be quite different in difficulty, etc. He found that after 20-40 items, the cumulative averages of all possible deletion patterns of cloze forms appear to stabilize. After 40 items, the difference between scores on easy and hard passages was well separated. However, for the hard passage, even after 175 items, there was still a range of difficulties from .25 to .43 among the five possible forms. At 50 items there was a spread of difficulties from .36 to .50 for the easy essay forms and from .20 to .54 over the hard forms.

2. Augmented Recall: Jenkinson.

Selected Processes and Difficulties of Reading Comprehension (1957). Jenkinson used the cloze procedure in producing protocols which were then used to study the verbal behavior involved in the act of filling in the blank. She used high school students, all of whom read and responded to three cloze passages. A few of the subjects were asked to recall their thinking on the original cloze task and to respond orally to a new cloze passage.

Of most interest to the present study is the manner in which Jenkinson constructed her cloze tests. She did not use an automatic deletion pattern, but rather had three restrictive rules: 1) "key words," whose meaning could be inferred from the remaining context, were deleted; 2) deletions

were then made in such a way that the number of deleted words in each of three word classes were equal, with the further requirement that 3) 90% of the deletions were included in the first two thousand of Thorndike's word list and all deleted words must be among the first five thousand words on the list. Note that this deletion procedure is not automatic. Of all the theses reported, this is the only one which did not use an automatic or semi-random deletion process.

Jenkinson reports the following coefficients between the total cloze scores of her subjects and the indicated variables:

<u>Cooperative Reading Test, C-2: Vocabulary</u>	.78
<u>Cooperative Reading Test, C-2: Level of Comprehension</u>	.73
Comprehension questions over passages used in cloze tests	.82
Intelligence	.69

Jenkinson's analysis of items by word class was limited to determining item difficulties, so it is not possible to assess which items (or word classes) are contributing to the coefficients. The item difficulty analysis confirmed Taylor's findings that content words are more difficult to fill in than function words. A third class composed of adjectives and prepositions was of intermediate difficulty. It should be noted, however, that the comprehension test over

the passages was done immediately after the cloze task and students were required to use the cloze passage. They did not have a complete passage to work from.

Jenkinson found that significant cloze score differences across grade levels were paralleled by similar differences across other reading tests. Where the cloze score differences were not significant, neither were the differences across the other tests. The rest of her thesis is not germane to the present study.

3. Reading Comprehension: Rankin.

An Evaluation of the Cloze Procedure as a Technique for Measuring Reading Comprehension (1958). Rankin used cloze scores to assess gains in reading skill developed during a reading improvement course. He used college students who were taking reading efficiency classes. They were, however, representative of university norms for high school standing and grade point average.

His subjects took a number of tests, some as a regular part of the reading classes or freshman entrance exams and some especially for his study. Standardized tests taken include the Diagnostic Reading Test (1948) and the SA-S: Senior Scale (1957), a personality inventory providing scores for sociability and emotional stability. Five cloze tests were developed for his study. One was an any-word, every

5th word deleted type with a total of 134 deletions. The other four tests limited deletions to nouns and verbs, deleting alternate eligible words from the passage. These tests each had 44 deletions.

Rankin proposed many specific hypotheses involving the validity of the cloze procedure (and the deletion variations). Of special interest to this study are the following findings. Table 1 presents the coefficients for the cloze tests and the sections of the Diagnostic Reading Test.

Table 1

Relationship between Scores on Cloze Tests with
Two Deletion Patterns and Scores on Sub-Tests of
the Diagnostic Reading Test

(after Rankin, 1958)

Deletion Pattern	Diagnostic Reading Test			
	Content	Function (Structural)		Reliability (Guttman L4)
	Components	Components		
	Story Comprehension	Vocabu- lary	Paragraph Comprehension	
Any-word	.29	.68	.60	.84
Noun-verb	.40	.32	.29	.69

Table 1 supports Rankin's notion that the vocabulary and paragraph sub-sections tap the function (structural) components of a passage while the story comprehension section taps the content components of the passage. He had predicted

that the any-word cloze test would provide results correlating more highly with measures of structural components than would a noun-verb deletion pattern. The observed correlations confirm this prediction. Rankin concluded that the any-word deletion is a measure of structural comprehension or intelligence while the noun-verb deletion system is primarily a measure of lexical comprehension.

Table 2 presents the results of cloze tests done with and without the opportunity to read the original passage prior to taking the test. Table 2 is interpreted to mean

Table 2

Relationships between Pre-reading and Post-reading Cloze Test Scores (Noun-Verb Deletions) and Scores on the Diagnostic Reading Test

(after Rankin, 1958)

Administration	<u>Diagnostic Reading Test</u>			Reliability (Guttman L4)
	Story Comprehension	Vocabu- lary	Paragraph Comprehension	
Pre-reading	.40	.32	.29	.69
Post-reading	.47	.36	.46	.74
Difference	NS	NS	p = < .04	

that the components of a cloze score are not very different whether you have read the original passage or not. This table does not show gains across cloze tests attributable to having

read the original passage between testings. It is a comparison only of cloze scores earned before and scores earned after reading the passage. The paragraph comprehension section is the only one with a significant difference. It was suggested that the post-reading cloze task and the paragraph test are more alike than any other combination. In both cases the reader has read the original passage or paragraph and is able to return to the material while doing the test.

Rankin investigated the effect of allowing credit for synonyms in scoring the cloze tests. The subjective (synonyms) scoring did not improve the validity of the cloze test over objective (exact replacement) scoring. There were correlations from .86 to .92 between the results of the two scoring systems. He concluded that use of synonym scoring is not warranted because of the increase in scoring time and loss of scoring objectivity required.

Gain scores between cloze forms when the original passage had been interpolated were assessed. Rankin found that there was a significant gain in cloze scores after reading when the same cloze test was used at both times. Practice effect was determined by using a control group who took a pre-reading cloze test and a different post-reading cloze test. No difference in gain was found between the two groups. Thus the change in cloze scores is credited to the effect of reading

the passage and not to practice on a cloze type test. The gain demonstrated by both scoring systems was significant, and the difference between the gain scores due to scoring system used was not significant. Considering the high correlations between scoring systems and the lack of significant differences in learning gains across the scoring systems, Rankin concluded that "...the use of a pre-post reading difference score between two objectively [exact replacement] scored cloze tests is a valid measure of the apprehension of meaning as a result of reading." (p. 154)

The relationship of intelligence to both pre- and post-reading cloze tests was reduced through the use of the content word (noun-verb) deletion restriction; however, it remained a significant factor in the scores. With respect to gain scores, the advantage held by the more intelligent subjects is lost. The more intelligent subjects had higher pre- and post-reading scores, but the gain scores were not related to IQ test standings. Reading rate was also shown not to be a significant factor entering into the post-reading cloze scores for the content word deletion pattern forms.

The final area Rankin considered had to do with the usefulness of cloze tests for measuring changes in reading skills. Using post-reading cloze tests administered at the start and finish of a six week reading efficiency course, he

found a small but significant gain in scores. However, the difference was equal to the difference in difficulty of the alternate forms. Thus, no improvement in cloze performance was indicated. Rankin suggests that the results may be due either to a lack of gain by the reading classes or to a lack of sensitivity in the cloze tests used to the kind of gain made.

Using pre-post reading scores obtained at the start and finish of the reading classes, no significant difference was found. As with the post-reading scores, the fact that he also found no differences reflected in his criterion measure (story comprehension section of the Diagnostic Reading Test) makes conclusions about the lack of gain demonstration tenuous at best. The rest of Rankin's paper deals with personality variables and is not germane to the present study.

4. Verbal Abilities: Fletcher.

A Study of the Relationships between Ability to use Context as an Aid in Reading and Other Verbal Abilities (1959). Fletcher compared cloze test performances with performances on other tests of verbal skills. He used college freshmen in his study to relate context usage (cloze performance) to reading comprehension, reading rate and general verbal ability. Verbal abilities were measured by means of the Cooperative English Test: comprehension section; the Dvorak-Van Wageningen Diagnostic

Examination of Silent Reading Abilities; and the American Council on Education Psychological Examination for College Freshmen (ACE).

Fletcher prepared his cloze tests over 250 word passages using an any-word, every 5th word deletion pattern. The three passages used were all 5th-7th grade level of difficulty. For data analysis, the cloze tests were pooled to provide a single context usage estimate for each subject. He found correlations of .59 and .57 between the cloze scores and the two reading rates available. With the Cooperative English Test he found correlations of .55 with the level of comprehension scores (and cloze), while the cloze performances correlated .63 with the vocabulary section. The ACE general verbal ability estimates correlated .72 with the cloze tests. The average cloze score was 42% (of 150 items).

Reliabilities were reported as .87 (split-half) and .87 (Kuder-Richardson #20). The close correspondence between the reliability estimates was taken to indicate a high degree of inter-item consistency. Using a high-low 27% procedure, Fletcher found the item discrimination indices (r), reported in Table 3. The item discrimination index, r , represents the correlation of an item with all other items in the test. It is, thus, a measure of the degree to which an item is sampling whatever the rest of the test is sampling. Note that nearly

Table 3
Distribution of Item Discrimination Indices ($\underline{r}$)
(after Fletcher, 1959)

Item $\underline{r}$'s	Percentage of Items (n = 150)
Less than .25	.47
.26 - .39	.24
Above .40	.29

50% of the items have item $\underline{r}$'s of less than .25, and only about 30% of the item $\underline{r}$'s are higher than .40. These two limits are frequently used in assessing items for retention in test revision efforts. Under these limits nearly 50% of the items are not working well enough to be retained in a revised test.

The difficulty level for each item was computed using all subjects and is summarized in Table 4. The distribution in Table 4 looks rectangular; with smaller categories the distribution appears slightly bi-modal, with a slight pile-up occurring at either extreme.

Fletcher examined the performance of each cloze item in light of its grammatical function. Dividing the items into content and function words, he found that the content words

Table 4
Distribution of Item Difficulty Indices (p)
(after Fletcher, 1959)

Item p	Percentage of Items (n = 150)
0 - .25	.39
.26 - .75	.33
.76 +	.28

were more difficult than the function words, while the item r's tended to be well distributed across both word classes. (64% of the items were content words.)

When the items were examined in terms of usage frequency information from other word counting studies, Fletcher found that the most frequently used words were the least difficult to supply and yet these words of high frequency had a slight tendency to occur among the items that discriminated best.

Fletcher also looked at the pool of words supplied for each cloze item. He found that the cloze responses looked very much like those of other controlled-association studies, i.e., a small number of responses to each item, with one word given a large percentage of the time, the remaining words rapidly dropping off in frequency. Interestingly, not only

did the low scorers leave more items blank, but they also tended to supply more words which were classed as uncommon responses than did the high scorers.

5. Aural Closure: Weaver.

An Examination of Some Differences in Oral and Written Language Using the Cloze Procedure (1961). Weaver presented cloze passages for reading and for listening. He used college juniors and seniors and administered cloze tests both as reading tests and as listening tests. The aural group was divided into those who were given a single hearing of the materials and those who were allowed as many repetitions of the listening materials as they desired. For both the aural and written presentation modes, half of the materials used had been originally prepared for oral presentation and half had been prepared for written presentation. Cloze tests were constructed with differing class restrictions on the deletions: any-word, every 10th deleted, and noun-verb, every 10th deleted.

These experimental variations resulted in a 2x2x2 schedule of eight cloze tests (channel: reading/listening; presentation: written/oral; deletions: content/any-word). Each cloze test had 40 items. Six of Weaver's hypotheses had to do with differences due to the presentation channel or the original channel intended for the material's presentation. The seventh

had to do with differences due to the content or any-word deletion patterns.

The reliabilities of the eight cloze tests (Flanagan test) were: reading .90, .86, .84, .58; listening .80, .78, .76, .74.

Although the opportunity to listen more than once to a passage increased the cloze scores, Weaver found that the relationships between the single and multiple listening situations were nearly the same.

The main finding was that the any-word deletion pattern is significantly more clozable when read, while a content word deletion is conveyed equally well by listening or reading. Weaver suggested that the loss in predictability of structural meaning (any-word deletions) was due to the relatively large context needed to supply all the structural meaning, compared with the more limited context which supplied predictive information sufficient for most lexical cases. In the listening situation it is very difficult to analyze large contextual organizations. This results in the structural deletion pattern being more difficult when presented aurally than when visually presented. When the subject listens, the information sequence and presentation correlates (rate, intonation, etc.) are under the control of the speaker. When the subject reads, the correlates (rate, repetition, selection of order of message portions for examination or reexamination,

plus some graphemic clues such as punctuation) are under his control. Thus, in a listening situation, memory must be adequate to provide for a review. In order to retain the message, most people appear to "chunk" material into larger meaningful patterns. The expected result would be that delayed recall will not produce exact duplicates of large amounts of input, but rather will reproduce an equivalent coding. This chunked memory is subject to all the effects of leveling, sharpening and individual idiosyncracies and will produce less exact replacement in a cloze task. Weaver did not score his cloze tests for synonyms, which would have been a nice check on this memory vs. inspection explanation of the reading-listening differences for structural deletions.

6. Readability: Bormuth.

Cloze Tests as a Measure of Readability and Comprehension Ability (1962). Bormuth used the cloze procedure with children to check on the applicability of its use with subjects younger than those previously studied. He used a population of 150 children in grades 4, 5 and 6. He investigated the relation of cloze performances and seven skills reported to be components of comprehension: 1) vocabulary in context, 2) stated facts, 3) sequence of ideas, 4) causal relationships, 5) main ideas, 6) inferences from facts, and 7) author's purpose.

He constructed an any-word, every 5th word deleted pattern over nine passages. The nine passages consisted of three difficulty levels in each of three topic areas - literature, social studies, and science. To check the comprehension variables he constructed 31 item, multiple choice tests over each of the passages with some questions intended to tap each of the seven components.

The 27 reliability coefficients reported (nine tests over three grades) ranged from .705 to .948 (split-half).

Validity coefficients between the cloze tests and the comprehension skills fills a 27/7 matrix. Of these, all but those in the main idea and the author's purpose columns exceeded the .01 level of significance. The main idea and author's purpose comprehension "tests" had rather low reliabilities, being based on only one question out of the 31 comprehension questions in each test. Of special interest are these correlations of the cloze results with the components: vocabulary .691 to .922; facts .584 to .870; sequence of ideas .364 to .751; and causal relationships .449 to .872. It should be pointed out that the reported coefficients are directly related to the number of questions in the comprehension tests tapping each comprehension variable, raising the question of restricted ranges affecting those tests with fewer items.

Testing the relative variances of the comprehension and cloze scores written on the same passages and administered to the same subjects, Bormuth concluded that both tests discriminate similarly between the grade level of subjects, difficulty of the passage, subject material covered, and their interactions. It was found that the cloze technique was more powerful in discriminating among passages differing in subject matter. The relation between the cloze score and the comprehension scores ranged from .73 to .84 for individual tests; when the scores were pooled for individual subjects, the relations increased to .82 to .95. The combining of the cloze and the comprehension tests was viewed by Bormuth as a means of lengthening the tests.

Table 5 gives the distribution of the item discrimination indices ($\underline{r}$) for Bormuth's cloze tests.

Table 5
Distribution of Item Discrimination Indices ($\underline{r}$)
(after Bormuth, 1962)

Item $\underline{r}$ ¹	Percentage of Items (n = 1350)
Less than .19	.21
.20 to .39	.20
Above .40	.59

¹It was not possible to use less than .25 and .26 to .39 ranges with the data as presented in the study report.

Table 6 shows the distribution of item difficulties.

Table 6
Distribution of Item Difficulty Indices (p)
(after Bormuth, 1962)

Item p	Percentage of Items (n = 1350)
.00 - .25	.43
.26 - .75	.42
.75 - 1.00	.15

7. Language Patterns: Ruddell.

An Investigation of the Effect of the Similarity of Oral and Written Language Structure on Reading Comprehension (1963). Ruddell constructed cloze tests over passages which had been constructed to control for linguistic patterning. He used 131 fourth graders to investigate the differences in comprehension that occur if written material is modified to approximate spoken language patterns. In this study, scoring the cloze tests was done both by the exact and the synonym methods. Ruddell constructed six cloze tests of controlled patterns of language structure by an any-word, every 5th word deletion pattern. Each cloze test had 50 items. Three tests were of high frequency of oral syntax

pattern and three had low frequency patterns.

Table 7 presents the split-half reliabilities for the scoring methods across the language patterns.

Table 7
Reliabilities (Split-half) of Cloze Tests
(after Ruddell, 1963)

Frequency of Oral Syntax Patterns in Written Passages	Scoring Methods			
	Exact		Synonym	
	Single tests	Pooled tests	Single tests	Pooled tests
High frequency passages	.851		.919	
	.874	.955	.891	.961
	.878		.899	
Low frequency passages	.866		.909	
	.896	.953	.913	.966
	.899		.916	

Average cloze scores were 31% (150 items) for high frequency passages and 26% for low frequency patterns when exact scoring was used. Synonym scoring raised the average scores to 38% and 34% respectively. Validity coefficients presented in Table 8 show little differences between scoring systems. Both scoring systems demonstrate highly significant relationships with all degrees of oral language frequency used in the written passages.

Table 8

Validity Coefficients of Cloze Tests and the
Stanford Achievement Test, Paragraph Section.

(after Ruddell, 1963)

Frequency of Oral Syntax Patterns in Written Passages	Scoring Methods	
	Exact	Synonym
High frequency passages	.716	.722
Low frequency passages	.771	.775

Looking for correlates of cloze performances, Ruddell found significant differences among children whose: fathers' occupations differed; mothers' educational background differed; intelligence scores differed; mental ages differed; and chronological ages differed. No significant difference in cloze scores could be found for sex differences. Note should be taken of the fact that the F ratios used in checking these correlates were similar in every case for the exact and synonym scoring systems over the background variables, but that great differences existed between the scoring methods across high and low language frequency patterns. The exact scoring method produced consistently higher values than the synonym method. Since there was little difference between the validities and reliabilities of the scoring methods, it

could be expected that the F ratios would also differ little. However, when scored exactly, the frequency of occurrence of language patterns accounted for significantly more of the variance of the cloze scores. Ruddell concludes that the results of tests scored by synonyms are affected by some variable or variables which do not enter into the exact scoring results.

D. Measurement Considerations

1. Prior closure tests.

Humans are able to fill in missing parts of a stimulus array or to organize complex stimulus arrays into a definite order. This ability has been used in studying many aspects of human behavior. There is a whole branch of experimental psychology which devotes itself to investigating the speed, flexibility, and stimulus conditions of perceptual closure. The organization of ambiguous arrays has been widely used in studying personality under the title of projective tests.

Closure has been used as well to assess specific knowledge and that class of general knowledge usually referred to as intelligence. The well known short answer and sentence completion tests both use the closure phenomenon. The Minkus completion test is a section of the Stanford-Binet Intelligence Test. (The Minkus requires the subject to supply function words.)

In the form of sentence completion tests, quite a number of current tests designed to assess reading skills use the closure task. One of these, Thurstone's Understanding Communication (1959) is used in the present study and is described in Chapter II.

The cloze procedure differs from these earlier uses of closure, in that the cloze deletions are distributed in a continuing passage and the cloze performance may depend on sensitivity to words and relationships distantly removed from the missing word.

2. Comprehension as process or product.

The term "comprehension" has at least two meanings. It can refer to the knowledge about a passage a person has at some given time. Comprehension can also refer to the process one goes through or is going through to gain knowledge from a reading passage. In the introduction to the problems of measuring reading comprehension, comprehension has been referred to as an ongoing process during reading by which the reader responds to the message contained in the passage. In much of the testing for comprehension, however, comprehension is normally treated as if it were a product. The problem appears to be one of developing methods which will allow assessment of the process. Jenkinson (1957) used the approach

of aided recall and of verbalization of ideas while solving a comprehension test. In addition to such intense study of an individual's behavior, means are needed to reduce the effect of those conditions which currently affect comprehension scores, but which may not be intrinsic to the reading process itself.

The usual reading comprehension test has inherently several conditions which introduce error and place the emphasis on comprehension as a product. Bormuth (1962) pointed out that imperfect comprehension by the test constructor limits the validity of the test as a measure of the reader's comprehension. Furthermore, lack of perfect communication by the examiner in his test items may limit the readers' comprehension of the test items. Finally, imperfect comprehension of the test items by the reader will appear as imperfect comprehension of the reading passage. Rankin (1958) notes that the same basic procedures may be present in the reading of everyone and yet the outcome (i.e., comprehension) may vary as a function of a number of conditions which affect usual test performances (i.e., prior knowledge, motivation, etc.) which are not intrinsic to the reading process. Since we wish to assess the reader's comprehension of the reading passage, variation of test item comprehension will confound our results. It may even be that the skills needed to

comprehend test items successfully are not the same as those needed to comprehend the original passage.

Use of the cloze procedure to measure comprehension appears to reduce the effect of these problems. To the extent that the process is random, the test constructors' comprehension, or lack of it, is irrelevant. With a cloze passage there are no other sources which need to be comprehended. Placing restrictions on the class(es) of eligible words appears to make it possible to control the effect of some factors normally involved in comprehension testing (intelligence, attitude, reading rate, etc.). If comprehension is viewed as a process, performance on a cloze test should reflect the reader's ability to follow what is being said. The cloze process does not ask what the reader can do with what he reads; it asks, can he read (i.e., apprehend the author's message)?

3. Factor analyses of comprehension.

Rankin (1962) reviewed a large number of factor studies of reading tests as well as earlier reviews of such studies. The factor studies seem to agree on four principal factors in the reading tests: a verbal factor (usually considered intelligence); a rate factor; a word knowledge factor; and a factor called knowledge of words in context.

By definition, factors are independent and, in practice, the independence of intelligence and reading rate can be demonstrated. Word knowledge and knowledge of words in context appear, therefore, to be logical components of reading comprehension. This study uses a power vocabulary test as an approximation to the word knowledge factor. A context vocabulary test (blanks in a passage, each with four choice words) is used to represent the words in context factor. Insofar as such tests measure the factors, they should be relatively independent of each other; if they are the key components of the comprehension act, they should together account for the variance of the total comprehension (cloze) scores.

4. Measuring vocabulary.

The word knowledge factor loads most highly on normal vocabulary tests. The power vocabulary test used in the present study follows a common format and is felt to be representative of this type of test.

The assessment of context vocabulary presents greater problems. When the ability to use the relationships between words is sought, the reader should know the message and simply be asked which word(s) carry the message best. In the cloze test, the question subsumes more than that. The reader must also determine the message. In this study, the problem of

providing the message is attacked in two ways. First, the passage on which the context vocabulary test was made is the same one as was originally read for the rate estimation section. Second, the choice of words for a particular space is provided by having four choice words given. As a check on the effect of having a complete context rather than one with deleted words, one group of subjects was given a context test with the key words underlined and numbered but not removed. The task was to pick the choice word that meant the same or nearly the same as the key word.

E. Problems in Using the Cloze Procedure

1. Length of cloze tests.

Problem: Taylor (1954) reported that after 20 to 40 items, the various possible test forms stabilized. Bormuth in a recent paper (1964) reported a study using the five possible forms of an any-word, every 5th word deletion pattern. He constructed all five forms (50 items) over 20 passages which covered five subject areas and ranged in difficulty from 4th to 8th grade level. He used large samples of children matched for IQ and reading achievement in grade four to eight.

Bormuth found that there were significant differences between the means of the different forms on over half the

20 tests even with a test length of 50 items. As the tests grew longer, the mean difficulties of the various forms showed less variance. However, the test length required to yield nearly identical mean difficulties may be longer than practical for most studies. For training, one form could be considered as good as another, but, for testing differences between groups or passages, it appears that one needs to consider the probability of obtaining the same results had he drawn a different pattern over the same material.

Conclusion: In this study two random (every 12th word) forms were drawn. If the means and variances of the two forms appear to be similar, it will be assumed that they represent the mean performances of all possible (12) forms. If they are dissimilar, then an estimate of the mean performance will need to be made. The test length of 50 items used in this study is expected to be sufficiently long to stabilize the results of the different forms.

2. Scoring methods.

Problem: The basic division in cloze scoring is in counting an item only if it matches exactly the deleted word or in giving credit also to synonyms which carry the meaning without distortion. Taylor (1953) reported on a series of small studies. From one of these studies, he reports the

results of scoring exact replacements only or of giving one-half credit for acceptable synonyms. He had three passages of differing difficulty and found that the proportion of the total score each passage accounted for was virtually identical under both scoring systems. The synonym scoring produced slightly larger scores, but there was no increase in ability to differentiate between passages.

Rankin (1958) scored his cloze test by exact and by synonym methods and found correlations between the two scores ranging from .86 to .92. In addition, the increase in scores arising from counting synonyms did not bring about a significant difference between gain scores when scored by both systems.

Ruddell (1963) used an elementary school population rather than an adult population to check on the effects of the different scoring procedures. He too found that both scoring methods gave significant results, but the advantage lay with exact scoring. Not only is exact scoring much easier to do, but it accounted for more of the variance of the cloze scores than did synonym scoring.

Conclusion: In the present study only exact scoring will be used.

3. Deletion restrictions.

Problem: Taylor (1957) reports the results of a three-way word-class restriction study. He used a chance deletion (every 10th word) over any-words, content words, and function words. He described content words as "hard" and function words as "easy" on the basis of an earlier study which found differences in the named directions when the word classes were used as cloze items. Taylor concluded that though all forms were significantly correlated with comprehension (multiple choice) tests over the same material, the function word deletions were much less effective. He chose the any-word over the hard word pattern for several reasons: simplicity of construction, usefulness for showing differences between materials (readability) as well as between individuals (comprehension), higher reliabilities, and higher correlation with intelligence.

Rankin (1958) used these data of Taylor's to recommend the hard word (content) pattern. Taylor had concluded that the any-word form was superior because the content form yielded only one correlation (with the test of pre-reading knowledge) higher than those yielded by the any-word pattern. To Rankin, the fact that the any-word form correlated as highly with intelligence as it did with pre- or post-reading tests indicated that it was measuring intelligence as much

as it was measuring comprehension. In contrast, the content word pattern predicted both pre- and post-reading comprehension better than it predicted intelligence. In addition to this reasoning, Rankin's study indicated that the any-word pattern is primarily a measure of intelligence, while a content (noun-verb) deletion restriction measures knowledge of the material of the passage.

Fletcher (1959) examined his any-word test by content and function items. He found the function words easier than the content words. However, he found function words more frequently among the items with high discrimination indices than were the content items.

Manis and Dawes (1961) looked at the "Cloze Score as a Function of Attitude." Two deletion rates (every 4th or every 8th word) did not produce significant differences. However, classification of the items as content or function words led to significant differences. The content blanks did not reflect differences due to the reader's previously expressed agreement or disagreement with the passage topic. The function blanks did show significant differences attributable to the reader's agreement with the passage stand. That is, the apparent effect of attitude is on function-word items.

Coleman and Blumenfeld (1963) used the cloze procedure to evaluate the effects of two styles of writing; active-verb and nominalized-verb sentences. The active verb sentences were easier to read. They had higher cloze scores, when the test had been constructed using any-word deletions. When the items were examined by class it was possible to locate the discriminating item group. The function items did not show any increase in score over the simpler to complex sentences, but the increase for content items was significant beyond .01.

Conclusion: In the present study, both any-word and content-word deletion patterns will be used.

4. Problems of direct observation.

Problem: Validity of a test has two components, reliability and relevance. Reliability can be measured in a number of ways without much difficulty. However, relevance, the correlation of a test with a criterion, becomes a problem when the criterion performance is not operationally definable. Reading rate is an example of a criterion performance that can be considered adequate. It can be defined, observed, and scored operationally. Reading comprehension, on the other hand, can not be observed, and so the best estimate of comprehension is the reader's score on some reading

test. When we use some such intermediate criterion, it must be as congruent with the behavior as logical analysis and research will allow, but we can never truly determine a test's validity. In the present study, this problem will be approached from two directions. First, the cloze test performance will be compared with the reader's performance on other tests thought also to measure reading comprehension. This approach really begs the entire question of what is the ultimate criterion. It merely asks, do two or more tests with similar labels tap the same skill(s)? This approach is what is normally called assessing concurrent validity.

Secondly, an attempt to analyze comprehension will be made using the results of factor analysis. Tests which are factorially simpler than comprehension will be used and the relation of these presumptive sub-skills to the total act of comprehension (cloze test performance) will be investigated. This is usually called construct validity. When a performance can be evaluated directly, as can rate of reading, then each item (in this case, each "word") is considered equal to every other. When the performance is scored for the quality of the act (e.g., a comprehension is right or wrong rather than completed) we still count the items, but we need also to check on the equality of the items. We must insure that performance on one item is equivalent to performance on another as an indicator of the total score. This

task will be done, in the present study, by examining the individual item effectiveness measures, inter-item r and item difficulty.

Conclusion: In the present study, the concurrent validity of the cloze procedure for assessing adult reading comprehension will be investigated through analysis of validity coefficients and item analysis results. Construct validity for the cloze procedure as a measure of reading comprehension will be developed through the attempt to predict comprehension (cloze) scores using the combined effect of a power vocabulary and a context vocabulary test.

Chapter II

Evaluation of Alternative Cloze Procedures for Assessing Adult Reading Comprehension

A. Development of Hypotheses

Hypothesis I: There is no difference in concurrent validity between the standard cloze procedure and the modified cloze procedure.

Discussion: Determination of validity involves consideration of both reliability and relevance. Validity coefficients of the cloze procedures with other tests will not tell the whole story by themselves. A series of questions to supply additional information about the validity of the cloze procedure and differences between procedures was formulated. The acceptance or rejection of this hypothesis will be based on evaluation of all information.

Question 1. Are there differences in the correlation coefficients between the cloze procedures and other reading tests used?

Discussion: Product moment correlation coefficients were used to assess the relationship of each of the cloze procedure types - stdC and modC (Michigan Adult Reading Test - constructed for this study) - with the Diagnostic Reading Test (Triggs, 1948), comprehension score and the Understanding Communication test (Thurstone, 1959). All coefficients were

expected to be positive and significant. The modification of the cloze procedure was expected to result in increased effectiveness of the cloze procedure. Significance of differences between the coefficients was determined to check the effect of the modification procedure.

Question 2. Are there significant differences between the two randomly deleted (stdC) cloze forms, Michigan Adult Reading Test (MART) forms B and C?

Discussion: The comprehension sections of MART forms A and D are identical by construction, so combining their populations is proper. MART forms B and C were both constructed by the stdC procedure, but since the deletion process began with different words, the two forms need to be compared for possible differences introduced by the chance variation of the alternative samplings over the same passage. It was felt that sampling two of the possible twelve forms would indicate the adequacy of either form by comparing the means and variances. If the two forms were to show significant differences, then further sampling of the remaining ten possible forms would be necessary to determine the mean difficulty and the variation from the mean or "passage" difficulty of each random sampling (form).

Question 3. Are there significant differences between the reliabilities of the stdC and modC procedures?

Discussion: The reliabilities of the MART subtests were computed by a split-half (odd-even) method with the Spearman-Brown correction applied, and with the Kuder-Richardson formula 21. The reliabilities of the Diagnostic Reading Test subtests and the Understanding Communication test were determined using the Kuder-Richardson 21. It was predicted that all reliabilities would be adequate, i.e., above .70. Furthermore, considering the rational deletion selection that went into the modC procedure, it was predicted that the reliabilities of the modC would be higher than those of the stdC.

Reliability of tests and sub-sections is critical in that it places upper bounds which limit correlation coefficients. It also restricts the value of a single score by an individual and the trust one can place in such a score for individual diagnosis.

Question 4. Are there any differences in the distributions of the item discrimination indices and item difficulties of the stdC and modC tests?

Discussion: An item discrimination index (item r) represents the correlation between each item and the variables which are being tapped by the rest of the test. Fan (1952) has published a table for item analysis which uses the differences in the proportions of correct responses to each

item found in the papers comprising the top and bottom 27% of the scores for each test. Using this table provides a quick means of estimating both the item discrimination and item difficulty indices.

The item discrimination indices (r) were summarized across the MART forms in the grouped frequencies of: r = less than .25; .26 - .39; and, .40 or better. This is a usual division for evaluating items when test revision is being considered. Such revision is not a direct part of this study, but the item "goodness" classification seems appropriate nonetheless.

Item difficulty is the second item statistic derived. This index can be computed by counting the number of correct responses for each item across all test papers and expressing it as the average probability of passing each item. However, the Fan table also provides an estimate of the average difficulty using the top and bottom 27% of the papers, and this estimate was read off at the same time as the item r was located.

Item difficulties (p) will be summarized in the difficulty ranges of: p = .00 - .05; .06 - .25; .26 - .75; .76 - .95; and, .96 - 1.00. This classification was based on the expectancy that when a random deletion procedure (stdC) is used there will be a considerable number of items

that will prove to be too easy or too hard, with fewer items spanning the mid-difficulty range. The rationally deleted items (modC) were expected, in contrast, to have few items at either extreme of difficulty, but rather to be more normally distributed about the mean difficulty - which was expected to be close to $p = .50$.

In selecting the high and low papers, forms A and D will be combined. Forms B and C consist of different items and will be analyzed separately. For the summary tables, the item analysis data will be summed across forms B and C to get a stdC performance estimate.

Item analysis information was also collected about the content and function word items in the stdC (any-word) forms. It will be used to supplement the information available about the differences between the stdC and modC procedures.

Question 5. What differences are reflected in the cloze scores that can be attributed to reading the original passage prior to taking the cloze test?

Discussion: To check on the benefit derived from reading the whole passage prior to responding to the mutilated cloze passage, some subjects attempted the cloze passages without first reading the original passage. The prediction was that pre-reading makes more difference to the success on the modC forms than on the stdC forms.

Hypothesis II: There is no difference in the construct validity estimates of the stdC and modC procedures.

Discussion: By theoretical postulation based on factor studies, comprehension has two components consisting of vocabulary power (words in isolation) and context vocabulary (words in context). The power vocabulary (Vp) and context vocabulary (Vc) tests of the MART were constructed to tap these two factors. Of four factors (rate and intelligence in addition to the two word factors) only intelligence has been controlled in this study. The range of intelligence was controlled by selecting only college upperclassmen and graduate students. The assignment to experimental groups, within this population, was random. Test questions to examine the other factors were devised.

Question 6. What is the relationship of reading rate estimates to the stdC and modC performances?

Discussion: Rate has proved to be a separate factor in factor analyses. Rankin (1961) has shown that when a significant relationship is demonstrated between rate and comprehension, it has been because the comprehension tests used were, in fact, speed rather than power tests. It was predicted that no significant relationship would be found between the rate estimates and the comprehension performances assessed by either the stdC or the modC procedures.

It would not be expected that the non-significant relationship would hold if either extreme of rate were restricted. Both extremely high and extremely low rates affect comprehension negatively. With normal college students, few have such excessive reading rates that they lose comprehension, but a rather considerable number of the students have such low rates that their comprehension suffers.

Question 7. Are there significant differences in the ability of the power vocabulary and context vocabulary tests jointly to predict the stdC versus the modC scores?

Discussion: The answer to this question was sought in the multiple correlation coefficients of the cloze tests by the joint contribution of the Vp and Vc tests. This multiple correlation (R) takes into account the relationships of each vocabulary test with the cloze performances and the relationship of the vocabulary tests to each other. Demonstration that both Vp and Vc have significant, positive correlations with the cloze scores, and that they (Vp and Vc) have a much lower correlation with each other would be viewed as support for the construct validity claim of the cloze comprehension procedure. Prediction was made that the R of both the stdC and the modC would be considerably higher than either of the vocabulary sub-tests alone.

MART form D differs from the other forms in that the key words in the context vocabulary section are underlined and numbered, but not removed from the passage. The key words and the distractor choices are constant for all forms; the difference is in whether there is a complete or mutilated passage present. Thus, there will be three multiple correlation coefficients to examine: the stdC; a modC group, the context vocabulary test having blanks; and, a modC group, the context vocabulary test having the words present.

B. Setting and Subjects

All of these data were gathered during the 1964 summer session at The University of Michigan. All subjects who took the full test battery were regularly enrolled School of Education students in one of three classes. The two groups that completed the cloze passages without having read the original passage were regularly enrolled students in English or Education.

The total population split 80% female/20% male and 56% undergraduate/44% graduate students. When the total population was divided into sample groups for data analysis, resulting proportions did not vary significantly from the above total proportions. For all groups, the testing session occurred during a regularly scheduled class period and all

testing of any group was completed within a single class session. The entire class populations were utilized; no prior information about the testing sessions, or chance to volunteer-or-not-participate was offered. The papers of two foreign students and four students who did not complete the test battery were eliminated prior to data analysis. All other subjects were used in the analysis. Two subjects who failed to note the number of the line they were reading during a reading rate task, but who had all other scores, were included. The final study population was 128.

C. Materials Used

Two types of cloze procedures were used to develop reading comprehension tests. The standard cloze procedure (stdC) consisted of deleting every 12th word from a 600 word passage. The modified procedure (modC) restricted possible deletions to content words (nouns, verbs, adverbs, and adjectives) and determined each deletion on the rational grounds of sufficiency of redundancy in the passage to allow exact replacement by a superior reader. Two widely used adult reading tests were used as markers against which to compare the results of the cloze tests. Each of the three tests will be described, and scoring procedures and resultant scores discussed. Samples of MARTtests are found in Appendix A.

The Michigan Adult Reading Test (MART) was constructed for this study. It consists of four sections, each of which provides a single score. Four forms (A, B, C, D) were developed. The forms are not parallel forms in the usual sense, but rather cover the same material using slightly different procedures. The differences in forms will be brought out during the description of each section. The time limits were generous enough so that each section can be considered a power rather than a speed test.

MART section 1: General Reading. This section consisted of a 2000 word reading passage adapted from chapter four, "The Empty Space Above," of The Blue of Capricorn by Eugene Burdick (1961). This section was identical for all forms of the test, and provided a measure of reading rate (M) expressed as words-per-minute. The testing procedure called for a notation of the line being read at the end of two minutes (upon the command "Mark") and at the end of five minutes (at "Stop"). At the end of five minutes all subjects went directly to section 2.

The passage was long enough and the time available short enough that only three subjects had finished the article prior to the stop order (e.g., read in excess of 400 w.p.m.). All subjects read fast enough to cover at least the first 600 words (e.g., above 120 w.p.m.). This

minimum coverage was desired so that all might be exposed to the material used in section 2, the first 600 words of the reading passage.

The double timing (Mark and Stop) was designed to ensure that unless a reader's rate was in excess of 1000 w.p.m., a rate could be obtained for him using the Mark record. The correlation between the Mark and Stop rates was .89. Since actual rates were available for every subject with the Mark recording, it was decided to use the Mark (M) rate for all data analysis.

MART section 2: Comprehension Check. This section consisted of the first 600 words of section 1, mutilated to produce 50 items by applying either a standard cloze deletion pattern (stdC) or a modified cloze deletion pattern (modC). The four forms provide three different cloze passages. MART forms A and D are identical in section 2 and were developed by the modC procedure. Forms B and C are both stdC passages differing only in where the deletion process was initiated.

The reader's task was to replace the deleted word. Twenty minutes were allowed for this section. Item credit was given only if the word supplied by the reader matched exactly the deleted word (obvious spelling errors excepted). The scores for the comprehension section are reported as

percentages of the total possible correctly completed. The success or failure of a single item would, therefore, make a difference of two points (2%) in the total score.

As a check on the benefit of prior reading on cloze performance, two classes (N = 60) took only MART section 2. Form A was used for one-half the group, and forms B and C were each taken by one-quarter of the group. The timing and scoring were identical with the procedure used for the full testing groups.

MART section 3: Power Vocabulary. This section was common to all four forms and consisted of 35 key words, each followed by four choice words. The task was to pick out the one choice word that meant the same or nearly the same as the key word. Ten minutes were allotted for this section. A single score (Vp) was derived, and represents the percentage of possible items correctly responded to. Thus, a single item represents about 3% of the total score.

The key words are not particularly esoteric. Rather, the power of the items lies in the close discriminations required between the choice words. This sample will illustrate the Vp format.

51. nourishment: 1) energy 2) diet 3) food
4) refreshment

The 35 key words were all used in the last 1400 words of the original passage (section 1). The basis for this

strategy of using one source for all the exercises was to maximize the familiarity of the reader with the material. Maximizing knowledge of the required task and familiarity with the format was also stressed in the instructions. The desire was to reduce the difference attributable to testing conditions by providing the subjects with maximum knowledge of procedural requirements and maximum material familiarity.

MART section 4: Context Vocabulary. In this section forms A, B, and C were identical. Form D used the same key words as the other, but left them in the passage rather than deleting the test words. The basic format for all forms was a passage of just over 500 words (which was edited and slightly revised from the last 1400 words of section 1). Of these 500 words, 35 were used as test words.

In forms A, B, and C, the 35 key words were deleted, replaced by numbered blanks, and accompanied by four choice words. The reader picked the one choice word which best fit into the passage at the indicated location. The key words were the same as those used in section 3; however, in this section, the key word was completely missing: deleted from the passage and not appearing as a choice word. The choice words were identical to those used for the power vocabulary test. This sample illustrates the Vc format. (Note the relationship of Vc 36 to Vp 51 used previously as an example.)

"Hawaiian storm petrels skim hungrily over the water taking some unknown 86 from the water...."

86. 1) energy 2) diet 3) food 4) refreshment

Form D differs in that the key word was underlined and numbered, but not deleted from the passage. Thus, form D used the same key words and choices as did the other Vc forms. The difference was in the presence of the key words in a meaningful reading passage. Vc item 86 appears in form D as:

"Hawaiian storm petrels skim hungrily over the water taking some unknown nourishment from the water...."
86

86. 1) energy 2) diet 3) food 4) refreshment

Fifteen minutes were allowed for this section. Scores were expressed as percentages of the total possible which were correctly responded to. A single item represents about 3% of the total score.

It was desired to have the context vocabulary test cover material where the content would not be a problem and the reader would be able to depend on the relations of the words to select the choice word. Using the original passage as a source for this section led to the comfortable feeling of familiarity. To place the burden of the decision on the relationships of the remaining words meant also that

the choice words should each fit into the context vocabulary blanks and the choice would be one of best fit rather than merely of sense. This demand for close distractors also fitted into the scheme used for the power vocabulary test. In the power test, the item difficulty was designed to arise not from esoteric key words but from close discriminations required between the choice words.

Diagnostic Reading Test (DRT), Survey Section, Upper Level, form A. The DRT is a widely used adult level reading test. All subjects in this study took a single form, A. As scored for the present study, the DRT provided three scores: reading rate (R); comprehension (TC), and, vocabulary (V).

DRT section 1: General Reading. This section consists of a 1200 word reading passage followed by twenty multiple choice comprehension questions. A three minute reading performance was used to compute the reader's rate. After the three minute marking of the line he was reading, the subject was instructed to finish the passage as quickly as possible. If he had not already done so, the reader was instructed to go directly to the questions at the end of eight minutes (five minutes after the three minute mark).

There was a total time limit of 15 minutes for this section, of which at least seven minutes were guaranteed

for the comprehension questions. Having once turned to the questions, the reader was not to return to the reading passage. The rate score (R) was expressed in words per minute. The comprehension questions were scored as the percentage of total possible items correctly marked. This comprehension score was added to the paragraph comprehension score of section 3 to produce the total comprehension score (TC) used in the data analysis of the present study.

DRT section 2: Vocabulary. This section consisted of 60 definitions, each followed by five choice words. The subject was to pick the one choice word that best fit the given definition. Ten minutes were allowed for this section. The scores were expressed as the percentage of attempted items correctly answered. This scoring differs from that indicated in the test manual where the vocabulary score is simply the number of correct items. However, even with the population restricted to college students, as in this study, not all subjects were able to attempt all items in the allotted ten minutes. To provide a score that reflects the test performance as a power rather than a speed score, the number of items attempted was used as a base. Most subjects attempted nearly all of the items (55 to 60), but the range extended down to only 35 items. Thus, the value of a single

vocabulary item ranges from about 1½% to 3% of the total score.

DRT section 3: Paragraph Comprehension. This section consisted of four short paragraphs, each of which was followed by five multiple choice questions over the content and implications of the paragraphs. The reader could refer back to the paragraphs while completing the questions about them. Fifteen minutes were allotted for this section.

Score 3 (comprehension) of the DRT manual consists of the combined story and paragraph comprehension scores. For this study the story and paragraph scores were averaged and called total comprehension (TC). The TC score was used in all data analysis of this study.

Understanding Communication. This test consisted of 40 short paragraphs, each having its last word deleted. Four choice words follow each paragraph and the reader was to choose the one word which best completed the paragraph. There was a fifteen minute limit to this test. The manual indicates that this should be sufficient to consider the test as a power rather than a speed test. The scores (UC) were expressed as the percentage of total possible items correctly completed. Each item is worth 2½% of the total score.

In her manual, Dr. Thurstone says, "Marking the right answer does not depend on memory or recognizing the answer

in the paragraph; rather, it depends on solving a problem presented in verbal form." (1959, p. 1) She subtitled this test "Verbal Reasoning." Though it looks more like the cloze test in format than does the DRT, to the extent that it really requires problem solving rather than understanding of what is being said in the paragraph, it should correlate higher with intelligence estimates than reading comprehension scores. With the alternate cloze forms used in this study, the prediction would be for a higher correlation of this UC test with the stdC than with the modC, since other studies have shown an any-word deletion pattern to be more highly related to intelligence scores than are content-word cloze tests.

Chapter III

Validity Analysis of Cloze Test Procedures

Hypothesis I: There is no difference in the concurrent validity between the standard cloze procedure and the modified cloze procedure.

It was assumed, on the basis of earlier studies, that the cloze procedure is a valid measure of adult reading comprehension. This hypothesis was intended to evaluate differences between cloze tests based upon differing deletion procedures.

1. Correlation coefficients between the standard and modified cloze forms and other reading tests will be equivalent.

Table 9 presents the coefficients relevant to question 1.

Table 9

Correlation Coefficients of the Michigan Adult Reading Test (MART) and Other Reading Tests

Standardized Reading Test	MART forms B + C (stdC, N = 65)	MART forms A + D (modC, N = 63)	Test	Signifi- cance
<u>Diagnostic Reading</u> <u>Test, Total Compre-</u> <u>hension</u>	.511 ¹	.670	z = 1.381	p > .05(NS)
<u>Understanding</u> <u>Communication</u>	.612	.520	z = 0.718	p > .05(NS)
Test	z = 0.838	z = 1.264		
Significance	p > .05(NS)	p > .05(NS)		

¹All coefficients significant at the .01 level.

(Complete data analyses appear in Appendix B.) None of the differences between correlation coefficients presented in Table 9 is significant. All of the correlations are significantly higher than chance, thus confirming the validity of the cloze procedure as a means of assessing reading comprehension. However, neither cloze task predicts the standardized comprehension tests significantly better than does the other. These coefficients indicate that the simpler procedure, automatic deletion of every nth word, yields results equivalent to the rational, non-random and word class restrictive cloze procedure used in this study, when other adult reading tests constitute the criterion.

Consideration should be given to the restricted ranges which were found with both the Diagnostic Reading Test (DRT) and the Understanding Communication (UC) tests. The mean percentage of the DRT-Total Comprehension (TC) was 79.2 ($s = 12$). The mean of the UC was 90.5 ($s = 4$). The MART seems more appropriate for the experimental population used since the stdC mean is 54.7 ($s = 10$) and the modC mean is 58.0 ($s = 15$). A less capable population or a more difficult test for this population might have yielded relationships masked by the restricted ceiling on these tests.

It is not possible to determine the contribution of the rational deletion procedure independent of the word class

restriction in the modC test. However, item analysis data (see 3 below) indicate that the deletion procedure affected the modC items beyond what might have been expected from the content word restriction alone.

2. The two MART forms, B and C (stdC) are equivalent despite different starting points for automatic deletion.

Table 10 presents the relevant tests. Neither the means nor the variances of the two forms are significantly different.

Table 10

Comparison of Two Forms (B and C) of the Michigan Adult Reading Test (MART) Constructed by the Standard Cloze Procedure

Central Tendency	MART		Test	Significance
	Form B (N = 32)	Form C (N = 33)		
Mean	55.1	54.1	t = 0.897	p .05 (NS)
Standard Deviation	11.0	9.0	F = 1.51	p .05 (NS)

Since the two stdC forms are equivalent, the results can be pooled and used for a single stdC population. Thus, except for item analysis, one stdC group will be used in the later data analyses.

Taylor (1954) stated that between 20 and 40 cloze items should stabilize alternate forms over the same passage.

Bormuth (1964) cautioned that even with 50 items, determinations of form equivalence should be made when using cloze tests for research. The stability of MART forms B and C may be due to the test length (50 items) or to the fact that only about 8% of the passage was removed by deletions, while both Taylor and Bormuth used 20% deletion rates.

3. The reliabilities of the stdC and modC procedures are equivalent.

Table 11 presents tests for assessing differences in reliability. The modC procedure is significantly more reliable

Table 11
Reliability Coefficients of Standard (stdC) and
Modified (modC) Cloze Tests

Reliability Estimate	stdC (N = 65)	modC (N = 63)	Test	Signifi- cance
Split-half + Spearman- Brown Correction	.523	.757	$z = 2.320$	$p < .05$
Kuder-Richardson #21	.760	.901	$z = 2.597$	$p < .01$

than the stdC procedure, as inferred from either reliability estimate. With the exception of the split-half estimate of the stdC test, all reliability estimates are above the .70 level which had been adopted as a criterion of adequacy.

4. The distribution of the item discrimination and item difficulty indices of the stdC and modC tests will be equivalent.

Table 12 presents the distributions of item discrimination indices ($\underline{r}$). While both the stdC and the modC procedures

Table 12

Item Discrimination Indices ($\underline{r}$) of Two Kinds of Cloze Test
(Percentages of Items)

Distribution $\underline{r} =$	stdC (n = 100)	modC (n = 50)	stdC Items by Grammatical Class	
			Content Items (n = 52)	Function Items (n = 48)
less than .25	.38	.14	.36	.40
.26 - .39	.07	.22	.06	.08
.40 or larger	.55	.64	.58	.52

result in nearly equal percentages of "good" or effective items ($\underline{r}$ above .40), there is a reversal of proportions across cloze procedures in the "poor" ($\underline{r}$ less than .25) and middle ranges. When the stdC items are grouped by word class, there is no difference between content and function items in the proportion of such items found in each discrimination range.

The modC procedure differs from the stdC in its use of a rational selection procedure and in its restriction of deletions to content words. It is clear that there is a much higher percentage of effective items in the modC test: 86% good and medium items for the modC procedure; 62% for the stdC. The question arises as to whether this difference is due to the selection process or to the word class restriction.

When the stdC tests are examined in terms of content and function word items, it becomes clear that the difference between the stdC and modC item discrimination indices is due to the selection process and not to the content word restriction. The distribution of proportions across the content-function dichotomy are equal. The rationally selected content words (modC) are distributed differently than are the randomly selected content words (stdC).

Table 13 presents the item difficulty distributions. The two cloze procedures result in similar distributions of item difficulty with one exception. The stdC form contains a substantial number of difficult items; 13% in the extreme difficulty range ($p = .00 - .05$). When the stdC items are grouped by word class, large differences in the difficulties between content and function words appear.

Table 13
Item Difficulty Indices (p) of Cloze Tests
(Percentages of Items)

p	stdC ($n = 100$)	modC ($n = 50$)	stdC Items by Grammatical Class	
			Content Items ($n = 52$)	Function Items ($n = 42$)
.00 - .05	.13	.00	.23	.02
.06 - .25	.09	.18	.13	.04
.26 - .75	.46	.46	.45	.46
.76 - .95	.27	.34	.17	.38
.96 - 1.00	.05	.02	.00	.10
Average p	.60	.56	.38	.71

Are the differences in item difficulties between the modC and the stdC due to the selection process or to the word class restriction? Looking at the difficulties of the content and function word items from the stdC leads to the conclusion that it was the selection process and not the word restriction which made the difference. Nearly a quarter of the randomly selected content words were extremely difficult ($p = .05$ or less), while none of them were extremely easy ($p = .96$ or more). The content items which were rationally selected, on the other hand, excluded all items of extreme difficulty and allowed only one item of extreme ease.

The obtained average difficulties of the content and function words, randomly deleted, confirm the findings of earlier studies, $p = .38$ and $.71$ respectively.

5. Differences in the cloze scores can be attributed to reading the original passage prior to taking the cloze test.

Table 14 presents the relevant data. None of the differences between the means is significant. The variance of the post-reading modC is significantly larger than either the pre-reading modC or the post-reading stdC.

Although both the stdC and the modC test results appear to be higher after reading, neither of these apparent gains is significant. This lack of significance is taken to indicate that the cloze process is indeed measuring the process of comprehension rather than the product. That is, the cloze procedure appears to be measuring information gained during reading rather than as a result of prior reading.

Summary. Analysis of all data indicates that the cloze procedure per se is a valid measure of adult reading comprehension. Data relative to test reliabilities and item analyses would suggest that the modC procedure is more effective than the stdC. Whatever the cloze procedure measures, the modC does more reliably and with greater item homogeneity

Table 14

Cloze Performances with and without Reading Original
Passage Prior to Taking Tests

Reading Condition	stdC (N = 30)	modC (N = 30)	Test	Significance
With Reading	M 54.7	58.0		$p > .05(\text{NS})$
	s 10.2	15.4	$F = 2.29$	$p < .01$
Without Reading	M 46.4	46.0	$t = 0.02$	$p > .05(\text{NS})$
	s 12.6	10.0	$F = 1.57$	$p > .05(\text{NS})$
Test	$t = 1.00$	$t = 1.29$	$F = 2.35$	
Significance	$p > .05(\text{NS})$	$p > .05(\text{NS})$	$p < .05$	

than does the stdC. Lack of significant differences between the concurrent validity coefficients of the two cloze procedures may be the result (1) of the restricted ranges of the standardized tests used, or (2) of the highly restricted population used, or (3) of insufficient refinement of grammatical classification and restriction. Thus, although the null statement of Hypothesis I cannot be clearly rejected, there is evidence that it should not be accepted with much confidence. Chapter IV indicates some future efforts that might be made to allow clearer acceptance or rejection of Hypothesis I.

Hypothesis II: The two vocabulary tests (Vp and Vc) will not account for the total variance of the cloze test results.

Based on the results of factor analyses, it was proposed that reading comprehension has two major components: power vocabulary and context vocabulary. Individual differences in intelligence were "controlled" in this study simply by truncating the IQ range, i.e., only college students were used. Whatever the actual range and distribution of intelligence within the experimental population, it is certainly more restricted than in a normal population of all adult readers. In addition to intelligence, rate was the other factor appearing in the factor studies.

6. Reading rate is unrelated to performance on the stdC and modC. Table 15 presents the reading rate relationships.

Table 15
Relationship of Reading Rate to Cloze Scores

Rate Estimate	stdC (N = 65)	modC (N = 63)	Test	Significance
MART, Mark Timing	.259*	.216*	$z = 0.027$	$p > .05$ (NS)
<u>Diagnostic Reading</u> <u>Test, Rate</u>	.469**	.109	$z = 2.197$	$p < .05$

*Significant at the .05 level.

**Significant at the .01 level.

Both cloze procedure tests are related significantly to reading rate. It was expected that no significant relationship between reading rate and cloze scores would be found. However, two of the four coefficients are significant at the .05 level and a third is significant at the .01 level. The modC procedure appears somewhat less affected by rate, but the question now arises as to whether the subjects did in fact have all the time they needed or not. If, as thought, sufficient time was provided, then speed would appear to be a dependent factor in comprehension as measured by the cloze procedures used here.

7. The power vocabulary and context vocabulary tests will not jointly predict cloze scores.

Table 16 presents the individual and joint correlations of the vocabulary and cloze tests. All correlations are significant at the .01 level. None of the differences between the coefficients reaches significance.

When blanks are used in the context vocabulary test (Vc), the correlation of the Vc to the Vp is of the same order as the correlations of the two vocabulary tests with both cloze procedure tests (.357 and .481). When the words are left in the Vc test, the correlation between the two vocabulary tests jumps to .797. This is interpreted as indicating that, with the key words in context, the context is little used, but rather the subject matches the imbedded key word with the choice words, yielding essentially a power test. Without the key word present, the subject uses the context to check the acceptability of the choice words. Chapter IV outlines a further investigation of this finding.

All of the correlations are significantly different from chance, indicating that the vocabulary tests, both power and context, share common elements with both cloze procedures. That none of the coefficients is significantly different across the cloze procedures is taken to indicate that the

Table 16

Relationships of Two Kinds of Vocabulary Tests and Two Cloze Tests

Vocabulary Test	stdC (Vc with blanks) (MART B + C, N = 65)	modC (Vc with blanks) (MART A, N = 33)	modC (Vc with words) (MART D, N = 30)	Test	Signifi- cance
Power Vocabulary (MART, Vp)	.356 ¹	.492	.492	$z = 0.884$	$p > .05(NS)$
Context Vocabulary (MART, Vc)	.432	.508	.592	$z = 0.445$ (blanks) $z = 0.949$ (words)	$p > .05(NS)$ $p > .05(NS)$
Multiple Correlation (R) of Power and Context Tests with Cloze Test	.484 ²	.648 ³	.592 ⁴	$z = 0.636$ (blanks) $z = 0.696$ (words)	$p > .05(NS)$ $p > .05(NS)$

1 All coefficients are significant at the .01 level.

2 r_{vpvc} (blanks) = .357

3 r_{vpvc} (blanks) = .481

4 r_{vpvc} (words) = .797

modification used in this study did not change the basic functioning of the cloze process.

The multiple correlations are not as high as was expected. The stdC has only about 23% of its variance accounted for by the two vocabulary tests. The modC has about 34% of its variance accounted for by using either of the context vocabulary formats. Reading rate has more in common with the cloze tests than was expected. Even so, there is a considerable amount of variance not accounted for in the cloze scores.

Summary. Hypothesis II cannot be rejected.

Chapter IV

Evaluation of the Cloze Procedure Modification Results

This study was designed to investigate two problems concerning the cloze procedure, both related to its validity as an assessment device for measuring adult reading comprehension. The first problem was the effect on concurrent validity of modifying the standard cloze deletion process. The second problem involved construct validity. It concerned the components of reading comprehension as measured by the cloze procedure.

Hypothesis I postulated that there is no difference between the concurrent validity of the two cloze procedures (stdC and modC) and the standardized reading tests used. Five questions were asked to elicit information which would aid in accepting or rejecting the hypothesis. The questions produced the following information:

1. No significant differences were found between the correlation coefficients of the two cloze forms with the standardized reading tests used. All coefficients were significant, confirming the assumption that the cloze procedure measures whatever it is that is also measured by other reading tests. The modification of the standard cloze procedure did not, however, produce any significant change in the degree of relationship between the tests.

2. The two cloze forms constructed using the standard deletion procedure did not differ in mean scores or test variance. The two standard forms differed only in the starting word for the deletion procedure.

3. The modified cloze procedure produced a significantly more reliable test than did the standard procedure.

4. Item discrimination (r) indices showed that the rational deletion plan used in the modified cloze procedure improved the "quality" of the test items. The standard cloze tests had many more items with low inter-item correlations (r less than .25). By analyzing the items from the standard form in terms of their grammatical class, it was shown that the differences in discrimination between the cloze procedures was due to the rational selection used in the modified cloze procedure and not to the word class restriction. The rationally selected items differ from the randomly selected items in the distribution of inter-item r 's; the randomly selected content word items do not differ from the randomly selected function word items.

Item difficulty (p) indices showed that the two cloze procedures had nearly the same distribution of item difficulties. A notable exception was the presence of a considerable number of extremely difficult items in the tests constructed by the standard deletion procedure. When the standard

cloze test items were grouped by grammatical class, it became clear that the rational selection procedure of the modified cloze test had eliminated the extremely difficult items. The content words which had been randomly selected under the standard procedure were much more difficult ($p = .38$) on the average than the content words which had been rationally selected by the modified deletion procedure (mean $p = .56$).

5. The cloze test scores were not significantly affected by prior reading of the original passage.

On the basis of this information, Hypothesis I could not be rejected. The findings indicate that the cloze procedure is a valid means for measuring reading comprehension, but they indicate also that the modification of the standard deletion procedure used in this study does not increase the concurrent validity of the standard cloze procedure as a measure of comprehension. One exception to this conclusion is the superiority of the modified form with regard to reliability.

Hypothesis II concerned the extent to which two vocabulary tests (power and context) account for the variance of the cloze test results. In null form, it stated that the two vocabulary tests do not account for the cloze test variance. Two questions were asked to collect supporting information.

6. Reading rate was found to be significantly related to cloze test performance. No relationship had been expected.

7. Multiple correlation coefficients were used to assess the predictive effectiveness of the two vocabulary tests used jointly. The criterion was the score on the cloze test. The obtained multiple correlations did not account for more than one third of the cloze test variance. No differences in predictive effectiveness were observed between the cloze tests developed using the standard or the modified cloze deletion procedures.

On the basis of these data, Hypothesis II could not be rejected. A large amount of the cloze test variance is not accounted for by the factors tapped in the power and context vocabulary tests used in this study. (The additional variance accounted for by the rate factor is very small and much of the cloze test variance remains unaccounted for even with rate added.)

The results of this study are important to the growing body of knowledge about the cloze procedure and its correlates. The use of the standard deletion procedure, simply removing every nth word (with or without word class restrictions), has been suggested by prior studies. In the present study, an

attempt was made to improve the effectiveness with which the cloze procedure assesses reading comprehension by modifying the deletion procedure. Each possible deletion was evaluated for its value as a test item by examining the redundancy remaining in the passage and selecting those items which it was felt could be replaced exactly by a superior reader.

The time and effort this modified selection procedure entailed was much greater than that required by the automatic selection of the standard cloze procedure. The modification of the standard procedure did not make a significant difference in the cloze test's effectiveness as a test of reading comprehension.

The study also demonstrated that the composition of comprehension, as assessed by the cloze procedure, is more complex than was thought. It cannot be accounted for by means of the two types of vocabulary tests and the factor of reading rate alone.

Some Unresolved Problems

The discovery that the two standard cloze tests were so similar in results raises the question of whether this is due to the test length or to the deletion rate used. Most of the earlier tests have used a 20% deletion rate; this study used an 8% rate. A study which systematically varies the deletion rate across materials of differing difficulty taken

by subjects of differing competencies might provide an answer to this question.

The effectiveness of the test formats for the context and the power vocabulary tests needs investigation. In the present study, the same passage, key words, and choice words were used for the construction of both types of vocabulary tests. A study using power and context vocabulary tests of equal difficulty, but differing components, should aid in determining the best process for developing context tests. Such a study might also show a different relationship with cloze test performance.

Other authors have suggested that even with an automatic deletion procedure, restriction of word classes eligible for deletion may be useful when measuring reading comprehension. The present study lends some support to this view. When the standard cloze test items are divided into content and function items, correlations for each word class can be made with the other test scores. The coefficients of the content word items were consistently higher with all tests than were those of the function word scores. All coefficients and differences between the correlations are of small magnitude, but the direction is clear and is worth further investigation with larger samples of subjects and items. (Table D in Appendix

B contains these coefficients.) The content and function scores were found to be unrelated in this study ($r = -.233$, $p .05$ (NS)). Thus, automatic deletion of content words should lead to the most useful cloze test for measuring adult reading comprehension.

Appendix A

M A R T

MART cover sheet, used in all forms
Michigan Adult Reading Test

by Frank P. Greene

Experimental
Edition
For Use Only
With
Permission

General Directions:

This test is designed to provide information about your reading skills. Each section measures some part of the complex activity of "reading." To insure useful results, you need to understand the directions for each part. Read the directions carefully and work the sample for each section. If you do not understand what you are to do, ask the examiner before that section is begun. You will be told when to start and stop each section.

Read each question carefully, but do not work too long on any one question. The score is simply the total number correct; there is no penalty for guessing. Work quickly and try not to leave any questions unanswered. If you have any extra time left over, you may use it to go back over that particular section, but you may not return to a section after the time for it is over.

Do not open the test booklet until told to do so.

This is a reusable test booklet. Do not mark on it in any way. Mark all answers on your answer sheet. For each section, make sure you know where to mark your answers before you begin that section. Make sure your name and today's date are on your answer sheet now.

Form

A

O

O

O

M A R T Section 1: GENERAL READING

Directions:

This is a test of your reading speed and comprehension. Read the selection as quickly as you can and still understand what you read. When you finish reading, you will be asked to answer some questions on the material you have read.

After you have been reading for some time, "Mark" will be called. Note the number of the line you are reading and put it in the space labelled "Mark" on the answer sheet. (Locate this space on your answer sheet now.) After recording the line number, go right on and finish the article as quickly as you can.

After you have been reading for several more minutes, "Stop" will be called. When "Stop" is called, again note the number of the line you are on and record this number in the space labelled "Stop" on the answer sheet. Then wait until you are told to go on.

If you have any questions about the directions, ask them now; if not, wait for the signal to turn the page and begin reading. Do not start until told to do so.

THE EMPTY SPACE ABOVE*

1 The surface voyager in the Pacific has only three things
2 to see: the interior of his ship, the skin of the ocean and
3 the empty space above him. The shipboard view will depend
4 entirely upon the ship that carries him. Occasionally he will
5 be reminded that there is a life below the surface: the wake
6 of a porpoise, the fin of a shark, the skittering of a flying-
7 fish.

8 The third dimension, the emptiness above him, is a puzzling
9 thing. Polynesians sail with eyes darting from the ocean's
10 surface to the sky and back. Westerners travel with their eyes
11 on the surface. For the land dweller it may be inevitable, he
12 must always watch where his feet land. It is a primitive impulse
13 which even the most sophisticated cannot resist. This is sad,
14 for the spaces above the Pacific teem with activity.

15 Take the birds. For some reason they are invisible to
16 Westerners. Often the only bird noticed is the gull. It lives
17 at the edge of the Pacific, rarely penetrating the vast central
18 spaces. The most beautifully proportioned of birds, it is,
19 oddly, the hyena of the sea; it scavenges. It will ride the air
20 current made by the wind hitting the side of a ship. For hours
21 it will hover without moving a feather. Sometimes it comes so
22 close that one can see a single unblinking eye . . . and watch
23 the hard leathery lid come down over the eye. Its proportions
24 and its soar are flawless. When garbage is dumped it will start

*Adapted with the author's permission from Chapter Four, "The Empty Space Above," of THE BLUE OF CAPRICORN, by Eugene Burdick. Boston. Houghton Mifflin, 1961.

25 a long slow outward spiral, its wings still motionless, and
26 move for the first time only when it has to brake itself and
27 settle into the debris. Then its beak begins to peck with a
28 strange disturbing anxiety.

29 But beyond the gull the sky teems with birds. The Polynesians
30 have learned to see the birds for two reasons: they are food and
31 they are an aid to navigation. Because they are doomed to fly
32 regular routes, they are like pointers in the sky. For example,
33 the golden plover led the early Polynesians to Hawaii. The
34 shining cuckoo is also alleged to have guided other argonauts
35 from the Solomons to New Zealand. Western eyes will only
36 see a beading of black dots in the middle distance, but the
37 Polynesian has learned to make out the different kinds of
38 birds.

39 The birds of the Pacific range from the magnificent to the
40 antic and even to the pathetic. The albatross, for example,
41 is huge and elegant and somewhat aloof. It flies at a far
42 distance and at a high altitude for the entire day. It hangs
43 at the very edge of visibility. At night it settles on the water
44 to feed on the small squids and fish that come to the surface.

45 One of the most absurd and satirized of birds is the gooney-
46 bird. It is, despite its reputation, a beautiful bird with a
47 pure white body and black wings. Seen singly, it is strikingly
48 handsome. But seen in swarms, they become a joke. They are a
49 victim of some internal mechanism which forces them to do
50 identical things in precisely identical ways until they die.
51 Let a building be built and when the gooneys return, those

52 that passed over that bit of land before will plunge helplessly
53 and at full speed into the building, breaking their necks and
54 falling to the ground in heaps. It is sad to see flights of
55 beautiful birds commit suicide with awful precision. The thuds
56 come with machine-gun rapidity and it is a stout heart that can
57 watch for long.

58 Hawaiian storm petrels will skim over the water, their feet
59 touching so that they seem to be walking on the water, taking
60 some unknown nourishment from the water. The few samples that
61 have been caught have had only a slimy substance and bits of
62 pumice stone in their stomach. They lead a solitary life and
63 will approach a ship only at night when they are attracted by
64 lights. These tiny birds of sooty brown flash through an aura
65 of a mainmast light, circle it for several minutes and then vanish.
66 There is something ascetic, shy and courageous about the storm
67 petrel. Like an old maiden aunt, it hesitates to make a direct
68 approach and lives on a Spartan diet, but cannot resist a
69 sudden gleam of light.

70 The bos'n bird is another which travelers will seldom see.
71 It is a graceful pure white bird and it flies hundreds of miles
72 over the ocean with a powerful flapping motion. Oddly enough
73 this graceful bird cannot stand on the ground and takes off
74 with a beating motion of the wings which quickly lifts it into
75 the air. When it lands it comes in with a bone-crushing thud,
76 the force of which is absorbed by thick breast feathers as well
77 as an air cushion under the skin. Its inability to stand also
78 means that it must mate in the air. The mating is not

79 promiscuous or casual, for the male and female join together
80 some weeks before the first breeding period and remain together
81 for life.

82 One of the most beautiful birds in the Pacific is the
83 man-o'-war hawk, with a wingspread of over seven feet. It is
84 a black bird with a metallic gloss of green and purple on its
85 feathers. It has tiny atrophied feet which are useless for
86 anything except perching which it has brought to a fine art.
87 It can perch on the smallest twig and sleep peacefully. The
88 bird cannot swim, run or even stand. As a result, it cannot
89 land on the water to fish as do other birds. The man-o'-war
90 hawk has brought soaring to its greatest perfection. It can
91 live almost indefinitely in the air, its great wings incredibly
92 sensitive to any thermal rise or fall, hoarding its energy as
93 it floats without the slightest exertion. I once watched from
94 land as a man-o'-war spied a surface fish and stalked it. It
95 was a miracle of tactical ingenuity. It dropped effortlessly
96 and soundlessly from a few hundred feet, carefully keeping its
97 shadow from falling over the fish.

98 Its useless feet have doomed it to a life which is terrifying
99 and mean. When it is a chick in its nest, it must be constantly
100 protected for others of its species will hover over the nesting
101 ground ready to pounce even on its own kind. They make merciless
102 raids on better equipped birds. In the evening they will flow
103 out to sea in a smooth rippling stream, hundreds of birds dipping
104 and falling, then soaring, each bird motionless yet the whole

105 band moving ahead. They never touch one another and are quite
106 soundless as they move out. At some point they intercept birds
107 coming back from a day's fishing and promptly begin a raid on
108 the full beaks and stomachs of the other birds. They will fall
109 like a plummet past a bird, snatch a fish from its mouth and soar
110 away--only to have another man-o'-war tear the fish from its jaws.
111 Occasionally they will seem to attack a gorged bird which, in
112 its fright, vomits up the fish it was carrying back for its
113 young. Fish drop like great black raindrops, but none hit the
114 water for each is caught and devoured before it can touch the
115 surface. Back and forth the deadly game goes until the hapless
116 have been entirely looted or the hawks are satisfied.

117 When they return to earth, the man-o'-war must land with
118 the greatest care. If they land on level ground, they will
119 not be able to take off. But even the slightest incline will
120 allow them to raise their vast wings, as the morning heat creates
121 fugitive and tiny thermal layers, and begin to soar. They land
122 in a gentle silent descent, which is actually a glittering-eyed
123 life-and-death calculation and maneuvering for advantageous
124 spots.

125 When viewed from an island or the deck of a ship birds seem
126 to lead an effortless life. In fact, they are caught up in a
127 taut circle of too many mouths for too little food. The margin
128 for error is slight. Golden plovers, for example, will fight one
129 another viciously for a bit of food, leaving a tiny circle of
130 feathers to mark their battles. In Hawaii early settlers brought
131 in foreign birds to "beautify" the forests and discovered that

132 the more energetic aliens quickly dominated the food supply.

133 Many of the old Hawaiian birds have now disappeared.

134 Although the battle for survival is incessant among the
135 wandering birds, some manage to survive and the species
136 continues. But those birds which have given up the fight and
137 retreated to the safety of lonely Pacific islands have had strange
138 and bitter things happen to them. Some birds, like the kiwi and
139 the dodo, lost the capacity to fly and were forever rooted to the
140 ground. The dodo suffered the worst of nature's penalties--
141 it utterly vanished. Other birds, stupefied by generations of
142 undangerous and languid life, lost their fear of men and landed
143 on the heads and arms of the first sailors they saw--and were
144 promptly strangled and eaten. The white man's ships also brought
145 the big black seagoing rat; an animal against which the disarmed
146 birds had no defense. The rats glutted on their compliant
147 victims and in some cases would decimate a whole clump of trees
148 of its bird population in a single day, eating only the heads
149 of the birds.

150 The great typhoon storms of the Pacific are also an enemy
151 of the bird. As the speed of the wind rises and the rain thickens
152 the storm finally reaches a point where the bird cannot fly
153 against it and is buffeted wildly from one layer of air to
154 another. Finally, exhausted, they begin to fall from the skies.
155 Sailors have reported seeing long stretches of water, after a
156 typhoon, covered with the freshly dead bodies of plovers,
157 storm petrels, shearwaters. On breeding islands such as Laysan
158 there will be thousands of bodies of young birds killed during

159 a storm; so thick that they must be scraped aside before vehicles
160 can pass.

161 The great skies of the Pacific are no limitless haven of
162 escape. Here too there is an orderliness which is intolerant
163 of mistakes, makes no allowances for individuals, ruthlessly
164 eliminates the weak and the slow. Most of it takes place out
165 of the sight of man, but it is there.

166 Except for supplying food to shipwrecked sailors, the
167 birds of the Pacific have only two direct economic uses. One
168 of these was the supplying of plumes for ladies' hats in the
169 early 1920's. The Japanese descended on isolated bird islands
170 and netted birds by the thousands. They also had the annoying
171 custom of eating the cache of food which American warships
172 periodically dropped off on the islands for castaways. With
173 the end of the plume fad, the Laysan albatross became freed
174 from one form of enemy although that enemy was to return a
175 generation later and destroy the albatross for more urgent
176 reasons.

177 The most valuable economic use of birds is guano. Because
178 of its fish diet the guano of sea birds is an excellent
179 fertilizer. The average bird will drop several dollars worth of
180 guano a year. The volume of guano deposited by the billions
181 of sea birds is staggering. Fortunately, most of it is washed
182 out to sea by heavy rains or whipped away by winds. If guano
183 were allowed to accumulate undisturbed, some islands such as
184 the Galapagos and Baker would, literally, rise thousands of
185 feet into the air and be natural resources of the first magnitude.

186 They would also be exceedingly unpleasant places to live.
187 On some islands the droppings of the birds have been
188 mixed with the broken-down coral sand and been converted into
189 a high grade of lime phosphate. Islands with such deposits
190 are methodically mined, great slashes cut into them by bulldozers
191 and dump trucks. Such islands provide the only truly hideous
192 landscapes in the Pacific.

END OF TIMED READING SELECTION

If you finish the reading selection before the examiner calls "Stop," put the number 290 in the space labelled "Stop" on the answer sheet and then wait for the examiner to tell you to go on.

O

O

1

MART section 2, used in forms A and D (modC)
M A R T Section 2: Comprehension Check.

Directions:

This test consists of a selection from the passage you have just read. At various points in the selection, single words have been left out and replaced by numbered blanks. Your task is to fill in each blank with the same word that was taken out.

Look at this example.

A. Do not write in this test booklet; record all your answers on the _____ sheet.

Your job is to supply the missing word. On your answer sheet, "answer" has been filled in blank A because answer is the word which was left out of the sample sentence.

It generally helps if you will look quickly over the entire comprehension selection to get the overall idea, then work on the individual items. The words remaining around each blank can help you figure what the missing word must be.

There are fifty (50) items. You will have twenty (20) minutes to work on this section.

To record your answers, write the word you think of in the space on the answer sheet. The answer spaces are numbered from 1 to 50 to correspond to the numbers of the blanks in the selection. Try not to leave any questions unanswered. If you can not do one item, go on and come back to it later. The score will simply be the total of correct responses. There is no penalty for guessing, so you should try every item.

If you have any questions, ask them now; if not, wait for the signal to turn the page. Do not start until told to do so.

THE EMPTY SPACE ABOVE

The surface voyager in the Pacific has only three things to 1: the interior of his ship, the skin of the 2 and the empty space above him. The shipboard view will depend entirely upon the 3 that carries him. Occasionally he will be reminded that there is a life below the 4: the wake of a porpoise, the 5 of a shark, the skittering of a flying-fish.

The 6 dimension, the emptiness above him, is a puzzling thing. Polynesians sail with 7 darting from the ocean's 8 to the sky and back. 9 travel with their eyes on the surface. For the land dweller it may be 10, he must always watch where his 11 land. It is a primitive impulse which even the most 12 cannot resist. This is sad, for the spaces above the Pacific teem with activity.

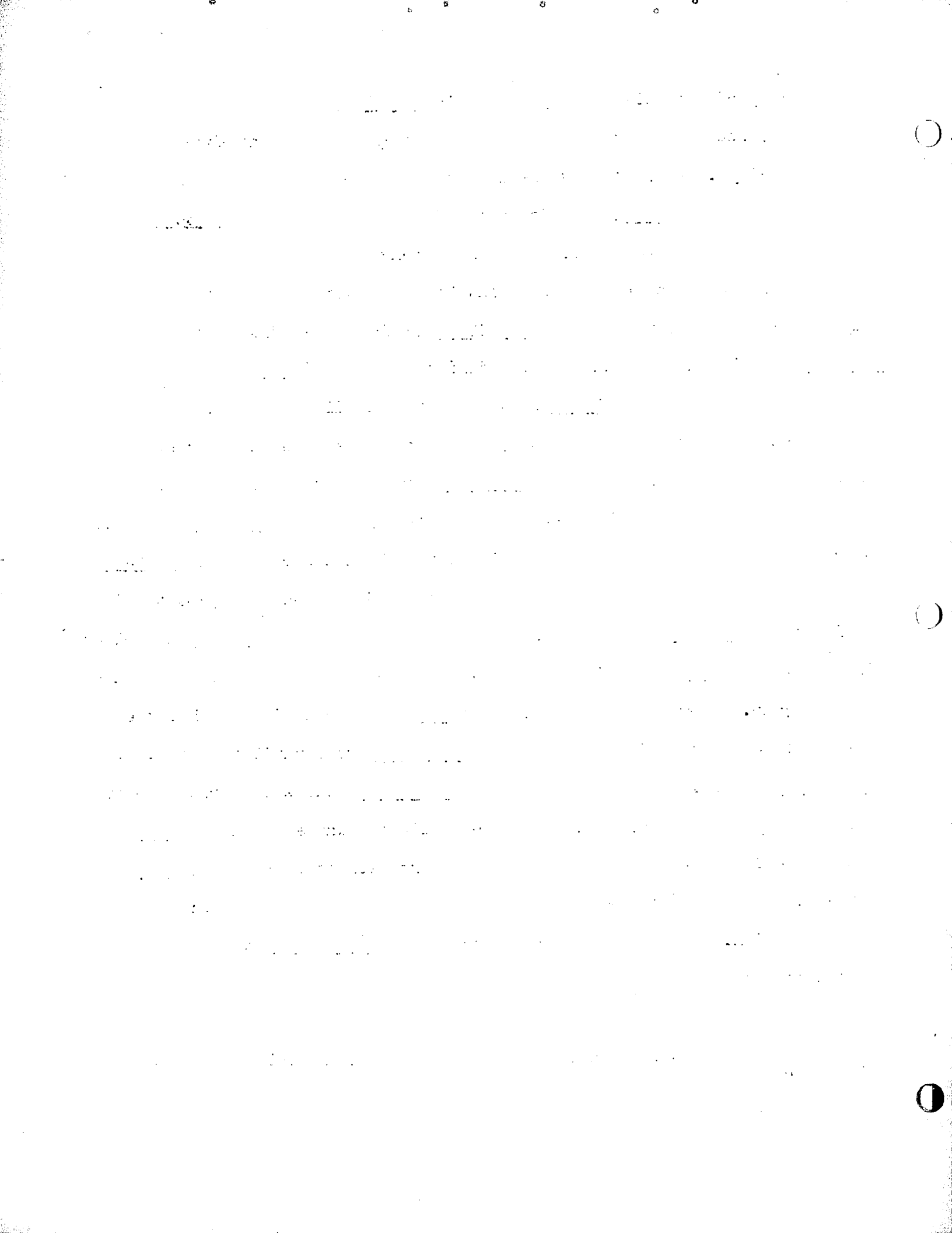
Take the 13. For some reason they are invisible to Westerners. Often the only bird noticed is the gull. It lives at the edge of the Pacific, rarely penetrating the vast 14 spaces. The most beautifully proportioned of 15, it is, oddly, the hyena of the sea; it 16. It will ride the air current made by the wind hitting the side of a 17. For hours it will 18 without moving a feather. Sometimes it comes so 19 that one can see a single unblinking 20 . . . and watch the hard leathery 21 come down over the eye. Its proportions and its soar are flawless. When 22 is dumped it will start a long slow outward spiral, its wings still 23, and move for the first time only when it has to brake itself and settle into the debris. Then its 24 begins to peck with a strange disturbing 25.

But beyond the gull the sky teems with 26. The Polynesians have 27 to see the birds for two reasons: they are food and they are an aid to 28. Because they are doomed to fly regular routes, they are like pointers in the 29. For example, the golden plover 30 the early Polynesians to Hawaii. The shining cuckoo is also alleged to have guided other argonauts from the Solomons to New Zealand. Western eyes will only 31 a beading of black 32 in the middle distance, but the 33 has learned to make out the different kinds of 34.

The birds of the 35 range from the magnificent to the antic and even to the pathetic. The albatross, for example, is huge and elegant and somewhat aloof. It flies at a far 36 and at a high altitude for the entire day. It hangs at the very edge of visibility. At 37 it settles on the water to feed on the small squids and fish that come to the 38.

One of the most absurd and satirized of birds is the gooney-bird. It is, despite its 39, a beautiful bird with a pure white 40 and black wings. Seen 41, it is strikingly handsome. But seen in swarms, they become a joke. They are a victim of some 42 mechanism which forces them to do identical things in precisely 43 ways until they die. Let a 44 be built and when the gooneys 45, those that passed over that bit of 46 before will plunge helplessly and at full 47 into the building, breaking their 48 and falling to the 49 in heaps. It is sad to see flights of beautiful birds commit suicide with awful precision. The thuds come with machine-gun 50 and it is a stout heart that can watch for long.

STOP HERE, UNTIL TOLD TO GO ON. IF YOU HAVE EXTRA TIME, YOU MAY GO BACK OVER THIS SECTION.



M A R T Section 3: Vocabulary Power

Directions:

This is a test of your vocabulary power. Look at the first word in each line. Then find another word in the same line that means the same or nearly the same as the first word. Note the number of this word and cross out the corresponding number on the answer sheet.

Look at this example.

B. Test: 1) repetition 2) blood 3) examination 4) investigation

On your answer sheet, number 3 has been crossed out because examination means the same or nearly the same as test.

There are thirty-five (35) vocabulary items. You will have ten (10) minutes to work on them. Your score will simply be the total number of correct items. There is no penalty for guessing. You should try every item. The items are not ordered by increasing difficulty. If you can not do one item, go on to the others and come back to it later. The answer spaces are numbered from 51 to 85 to correspond to the item numbers. Try not to leave any questions unanswered.

If you have any questions, ask them now; if not, wait for the signal to turn the page. Do not start until told to do so.

51. nourishment: 1) energy 2) diet 3) food 4) refreshment
52. slimy: 1) slippery 2) miry 3) greasy 4) sloppy
53. pumice: 1) powder 2) sandstone 3) volcanic-rock 4) pummel
54. solitary: 1) single 2) unhappy 3) abandoned 4) excluded
55. aura: 1) beam 2) aroma 3) glow 4) pinnacle
-
56. ascetic: 1) abstaining 2) intemperate 3) odd 4) sterile
57. atrophied: 1) dead 2) starved 3) atrocious 4) shrunken
58. perching: 1) sitting 2) climbing 3) standing 4) hanging
59. indefinitely: 1) securely 2) precisely 3) indifferently 4) indeterminately
60. incredibly: 1) uncertainly 2) plausibly 3) unbelievably 4) questionably
-
61. exertion: 1) effort 2) struggle 3) gain 4) exhaustion
62. spied: 1) glimpsed 2) searched 3) hid 4) led
63. stalked: 1) deserted 2) observed 3) waited 4) approached
64. tactical: 1) polite 2) concise 3) diplomatic 4) planned
65. vast: 1) infinite 2) great 3) indefinite 4) abundant
-
66. fugitive: 1) deceptive 2) elusive 3) permanent 4) trapped
67. thermal: 1) cool 2) vacuum 3) heat 4) temperature
68. descent: 1) slump 2) slide 3) drop 4) downcast
69. maneuvering: 1) attacking 2) scheming 3) defending 4) climbing
70. incessant: 1) continuous 2) frequent 3) random 4) immoral

71. species: 1)birds 2)breed 3)sample 4)items
72. retreated: 1)celebrated 2)withdrawn 3)recessed 4)proceeded
73. stupefied: 1)amazed 2)sensitized 3)shocked 4)dulled
74. languid: 1)fluent 2>alert 3)listless 4)energetic
75. gluttoned: 1)emptied 2)stuffed 3)gulped 4)devoured
-
76. decimate: 1)dismantle 2)decipher 3)devastate 4)demise
77. guano: 1)nest 2)excrement 3)food 4)carcass
78. deposited: 1)deposed 2)entrusted 3)saved 4)dropped
79. staggering: 1)overwhelming 2)overloaded 3)hesitating 4)dangling
80. accumulate: 1)receive 2)preserve 3)adjust 4)amass
-
81. resources: 1)revenues 2)means 3)assets 4)credits
82. magnitude: 1)pattern 2)power 3)dimension 4)height
83. methodically: 1)neatly 2)systematically 3)symmetrically 4)routinely
84. slashes: 1)caves 2)fractures 3)divisions 4)gashes
85. hideous: 1)ugly 2)mechanical 3)pungent 4)secret

STOP HERE, UNTIL TOLD TO GO ON. IF YOU HAVE EXTRA TIME, YOU MAY GO BACK OVER THIS SECTION.

O

O

O

M A R T Section 4: Context Vocabulary

Directions:

This is a test of your sensitivity to words in context. On the next pages, you will find a reading passage with certain words replaced by numbered blanks. On the facing pages, you will find the numbers of the blanks followed by four words. Pick the one word that best fits into the numbered blank in the passage. Note the number of this word and cross out the corresponding number on the answer sheet. Look at the sample

Look at the story and use the C to help you in picking your choice.

C: 1) blank 2) title 3) spaces 4) context

On your answer sheet, number 4 has been crossed out because "context" makes the best fit in the sample

There are thirty-five (35) vocabulary items. You will have fifteen (15) minutes to work on them. Your score will simply be the total number of correct items. There is no penalty for guessing. You should try every item. The items are not ordered by increasing difficulty. If you cannot do one item, go on to the others and come back to it later. The answer spaces are numbered from 86-120 to correspond to the numbers of the items. Try not to leave any question unanswered.

If you have any questions, ask them now; if not, wait for the signal to turn the page. Do not start until told to do so.

Hawaiian storm petrels skim hungrily over the water taking some unknown 86 from the water. The few samples that have been caught have had in their stomachs only a 87 substance from the sea and bits of 88 from some far off volcanic island or reef. They lead a 89 life and will approach a ship only at night when they are attracted by the lights. These tiny birds flash through the 90 of a mainmast light, circle it for several minutes and then vanish back into the darkness. There is something 91, shy, courageous and aloof about the storm petrel.

The man-o'-war hawk has a huge wingspread of over seven feet and tiny 92 feet which are useless for anything but 93 which it has brought to a fine art. It can perch on the smallest twig and sleep peacefully. It can live almost 94 in the air, its great wings 95 sensitive to any thermal rise or fall, hoarding its energy as it floats without the slightest 96. I once watched as a man-o'-war hawk 97 a surface fish and 98 it. It was a miracle of 99 ingenuity. It dropped effortlessly and soundlessly from a few hundred feet, carefully keeping its shadow from falling over the fish. If the men-o'-war land on level ground they will not be able to take off. But even the slightest incline will allow them to raise their 100 wings, as the morning sun creates 101 and tiny 102 layers, and begin to soar. They land in a gentle silent 103 which is actually a life-and-death calculation and 104 for advantageous spots.

86. 1) energy 2) diet 3) food 4) refreshment
87. 1) slippery 2) miry 3) greasy 4) sloppy
88. 1) powder 2) sandstone 3) volcanic-rock 4) pummel
89. 1) single 2) unhappy 3) abandoned 4) excluded
90. 1) beam 2) aroma 3) glow 4) pinnacle
-
91. 1) abstaining 2) intemperate 3) odd 4) sterile
92. 1) dead 2) starved 3) atrocious 4) shrunken
93. 1) sitting 2) climbing 3) standing 4) hanging
94. 1) securely 2) precisely 3) indifferently 4) indeterminately
95. 1) uncertainly 2) plausibly 3) unbelievably 4) questionably
-
96. 1) effort 2) struggle 3) gain 4) exhaustion
97. 1) glimpsed 2) searched 3) hid 4) led
98. 1) deserted 2) observed 3) waited 4) approached
99. 1) polite 2) concise 3) diplomatic 4) planned
100. 1) infinite 2) great 3) indefinite 4) abundant
-
101. 1) deceptive 2) elusive 3) permanent 4) trapped
102. 1) cool 2) vacuum 3) heat 4) temperature
103. 1) slump 2) slide 3) drop 4) downcast
104. 1) attacking 2) scheming 3) defending 4) climbing

GO RIGHT ON TO NEXT PAGE

Although the battle for survival is 105 among the wandering birds, some manage to survive and the 106 continues. But those birds which have given up the fight and 107 to the safety of lonely Pacific islands have had strange and bitter things happen to them. The Dodo suffered the worst of nature's penalties: it utterly vanished. Other birds, 108 by generations of undangerous and 109 life, lost their fear of men and were quickly caught and eaten. The white man also brought the black rat against which the dumb and slow birds had no defense. The rats 110 on their compliant victims and in some cases would 111 a whole clump of trees of its bird population in a single day.

Because of its fish diet the 112 of sea birds is an excellent fertilizer. The volume of guano 113 by the billions of birds is 114. Fortunately most of it is washed out to sea by heavy rains. If the guano were allowed to 115 undisturbed some islands would rise thousands of feet into the air and be natural 116 of the first 117. On some islands the droppings of birds have been mixed with the broken-down coral sand and been converted into a high grade of lime phosphate. Islands with such deposits are 118 mined, great 119 cut into them by bulldozers and dump trucks. Such islands provide the only truly 120 landscapes in the Pacific.

105. 1) continuous 2) frequent 3) random 4) immoral
106. 1) birds 2) breed 3) sample 4) items
107. 1) celebrated 2) withdrawn 3) recessed 4) proceeded
108. 1) amazed 2) sensitized 3) shocked 4) dulled
109. 1) fluent 2) alert 3) listless 4) energetic
110. 1) emptied 2) stuffed 3) gulped 4) devoured
-
111. 1) dismantle 2) decipher 3) devastate 4) demise
112. 1) nest 2) excrement 3) food 4) carcass
113. 1) deposited 2) entrusted 3) saved 4) dropped
114. 1) overwhelming 2) overloaded 3) hesitating 4) dangling
115. 1) receive 2) preserve 3) adjust 4) amass
-
116. 1) revenues 2) means 3) assets 4) credits
117. 1) pattern 2) power 3) dimension 4) height
118. 1) neatly 2) systematically 3) symmetrically 4) routinely
119. 1) caves 2) fractures 3) divisions 4) gashes
120. 1) ugly 2) mechanical 3) pungent 4) secret

STOP HERE, UNTIL TOLD TO GO ON. IF YOU HAVE EXTRA TIME, YOU MAY GO BACK OVER THIS SECTION.

Directions:

This test consists of a selection from the passage you have just read. At various points in the selection, single words have been left out and replaced by numbered blanks. Your task is to fill in each blank with the same word that was taken out.

Look at this example.

- A. Do not write in this test booklet; record all your answers on the _____ sheet.

Your job is to supply the missing word. On your answer sheet, "answer" has been filled in blank A because answer is the word which was left out of the sample sentence.

It generally helps if you will look quickly over the entire comprehension selection to get the overall idea, then work on the individual items. The words remaining around each blank can help you figure what the missing word must be.

There are fifty (50) items. You will have twenty (20) minutes to work on this section.

To record your answers, write the word you think of in the space on the answer sheet. The answer spaces are numbered from 1 to 50 to correspond to the numbers of the blanks in the selection. Try not to leave any questions unanswered. If you can not do one item, go on and come back to it later. The score will simply be the total of correct responses. There is no penalty for guessing, so you should try every item.

If you have any questions, ask them now; if not, wait for the signal to turn the page. Do not start until told to do so.

The surface ____1____ in the Pacific has only three things to see: the interior ____2____ his ship, the skin of the ocean and the empty space ____3____ him. The shipboard view will depend entirely upon the ship that ____4____ him. Occasionally he will be reminded that there is a life ____5____ the surface: the wake of a porpoise, the fin of a ____6____, the skittering of a flying-fish.

The third dimension, the emptiness above ____7____, is a puzzling thing. Polynesians sail with eyes darting from the ocean's ____8____ to the sky and back. Westerners travel with their eyes ____9____ the surface. For the land dweller it may be inevitable, he ____10____ always watch where his feet land. It is a primitive impulse ____11____ even the most sophisticated cannot resist. This is sad, for the ____12____ above the Pacific teem with activity.

Take the birds. For some ____13____ they are invisible to Westerners. Often the only bird noticed is ____14____ gull. It lives at the edge of the Pacific, rarely penetrating ____15____ vast central spaces. The most beautifully proportioned of birds, it is, ____16____, the hyena of the sea; it scavenges. It will ride the ____17____ current made by the wind hitting the side of a ship. ____18____ hours it will hover without moving a feather. Sometimes it comes ____19____ close that one can see a single unblinking eye. . . and watch ____20____ hard leathery lid come down over the eye. Its proportions and ____21____ soar are flawless. When garbage is dumped it will start a ____22____ slow outward spiral, its wings still motionless, and move for the ____23____ time only when it has to brake itself and settle into the ____24____. Then its beak begins to peck with a strange disturbing ____25____.

But beyond the gull the sky teems with birds. The Polynesians
26 learned to see the birds for two reasons: they are food
27 they are an aid to navigation. Because they are doomed to
28 regular routes, they are like pointers in the sky. For example,
29 golden plover led the early Polynesians to Hawaii. The shining
cuckoo 30 also alleged to have guided other argonauts from the
Solomons to 31 Zealand. Western eyes will only see a beading of
black 32 in the middle distance, but the Polynesian has learned to
make 33 the different kinds of birds.

The birds of the Pacific range 34 the magnificent to the antic
and even to the pathetic. The 35, for example, is huge and elegant
and somewhat aloof. It flies 36 a far distance and at a high
altitude for the entire 37. It hangs at the very edge of visibility.
At night it 38 on the water to feed on the small squids and fish
39 come to the surface.

One of the most absurd and satirized 40 birds is the gooney-
bird. It is, despite its reputation, a beautiful 41 with a pure
white body and black wings. Seen singly, it 42 strikingly handsome.
But seen in swarms, they become a joke. They 43 a victim of some
internal mechanism which forces them to do 44 things in precisely
identical ways until they die. Let a building 45 built and when
the gooneys return, those that passed over that 46 of land before
will plunge helplessly and at full speed into 47 building, breaking
their necks and falling to the ground in heaps. 48 is sad to see
flights of beautiful birds commit suicide with 49 precision. The
thuds come with machine-gun rapidity and it is a 50 heart that can
watch for long.

STOP HERE, UNTIL TOLD TO GO ON. IF YOU HAVE EXTRA TIME, YOU MAY GO BACK
OVER THIS SECTION.

M A R T MART section 2, used in form C (stdC)
Section 2: Comprehension Check.

Directions:

This test consists of a selection from the passage you have just read. At various points in the selection, single words have been left out and replaced by numbered blanks. Your task is to fill in each blank with the same word that was taken out.

Look at this example.

A. Do not write in this test booklet; record all your answers on the _____ sheet.

Your job is to supply the missing word. On your answer sheet, "answer" has been filled in blank A because answer is the word which was left out of the sample sentence.

It generally helps if you will look quickly over the entire comprehension selection to get the overall idea, then work on the individual items. The words remaining around each blank can help you figure what the missing word must be.

There are fifty (50) items. You will have twenty (20) minutes to work on this section.

To record your answers, write the word you think of in the space on the answer sheet. The answer spaces are numbered from 1 to 50 to correspond to the numbers of the blanks in the selection. Try not to leave any questions unanswered. If you can not do one item, go on and come back to it later. The score will simply be the total of correct responses. There is no penalty for guessing, so you should try every item.

If you have any questions, ask them now; if not, wait for the signal to turn the page. Do not start until told to do so.

THE EMPTY SPACE ABOVE

The surface voyager in the 1 has only three things to see: the interior of his ship, 2 skin of the ocean and the empty space above him. The 3 view will depend entirely upon the ship that carries him. Occasionally 4 will be reminded that there is a life below the surface: 5 wake of a porpoise, the fin of a shark, the skittering 6 a flying-fish.

The third dimension, the emptiness above him, is a 7 thing. Polynesians sail with eyes darting from the ocean's surface to 8 sky and back. Westerners travel with their eyes on the surface. 9 the land dweller it may be inevitable, he must always watch 10 his feet land. It is a primitive impulse which even the 11 sophisticated cannot resist. This is sad, for the spaces above the 12 teem with activity.

Take the birds. For some reason they are 13 to Westerners. Often the only bird noticed is the gull. It 14 at the edge of the Pacific, rarely penetrating the vast central 15. The most beautifully proportioned of birds, it is, oddly, the hyena 16 the sea; it scavenges. It will ride the air current made 17 the wind hitting the side of a ship. For hours it 18 hover without moving a feather. Sometimes it comes so close that 19 can see a single unblinking eye . . . and watch the hard leathery 20 come down over the eye. Its proportions and its soar are 21. When garbage is dumped it will start a long slow outward 22, its wings still motionless, and move for the first time only 23 it has to brake itself and settle into the debris. Then 24 beak begins to peck with a strange disturbing anxiety.

But beyond 25 gull the sky teems with birds. The Polynesians have learned to 26 the birds for two reasons: they are food and they are 27 aid to navigation. Because they are doomed to fly regular routes, 28 are like pointers in the sky. For example, the golden plover 29 the early Polynesians to Hawaii. The shining cuckoo is also alleged 30 have guided other argonauts from the Solomons to New Zealand. Western 31 will only see a beading of black dots in the 32 distance, but the Polynesian has learned to make out the different 33 of birds.

The birds of the Pacific range from the magnificent 34 the antic and even to the pathetic. The albatross, for example, 35 huge and elegant and somewhat aloof. It flies at a far 36 and at a high altitude for the entire day. It hangs 37 the very edge of visibility. At night it settles on the 38 to feed on the small squids and fish that come to 39 surface.

One of the most absurd and satirized of birds is 40 gooney-bird. It is, despite its reputation, a beautiful bird with a 41 white body and black wings. Seen singly, it is strikingly handsome. 42 seen in swarms, they become a joke. They are a victim 43 some internal mechanism which forces them to do identical things in 44 identical ways until they die. Let a building be built and 45 the gooneys return, those that passed over that bit of land 46 will plunge helplessly and at full speed into the building, breaking 47 necks and falling to the ground in heaps. It is sad 48 see flights of beautiful birds commit suicide with awful precision. The 49 come with machine-gun rapidity and it is a stout heart that 50 watch for long.

STOP HERE, UNTIL TOLD TO GO ON. IF YOU HAVE EXTRA TIME, YOU MAY GO BACK OVER THIS SECTION.

○

○

○

M A R T Section 4: Context Vocabulary

Directions:

This is a test of your sensitivity to words in context. On the next pages, you will find a reading passage with certain words underlined and numbered. On the facing pages, you will find the numbers of the underlined words followed by four choice words. Pick the one choice word that means the same or nearly the same as the underlined word in the passage. Note the number of this word and cross out the corresponding number on the answer sheet. Look at the sample.

Look at the whole story and use the context to help you in making your word choice.
C

C: 1) blank 2) title 3) spaces 4) setting

On your answer sheet, number 4 has been crossed out because "setting" makes the best substitute for "context" in this sample.

There are thirty-five (35) vocabulary items. You will have fifteen (15) minutes to work on them. Your score will simply be the total number of correct items. There is no penalty for guessing. You should try every item. The items are not ordered by increasing difficulty. If you cannot do one item, go on to the others and come back to it later. The answer spaces are numbered from 86-120 to correspond to the numbers of the items. Try not to leave any question unanswered.

If you have any questions, ask them now; if not, wait for the signal to turn the page. Do not start until told to do so.

Hawaiian storm petrels skim hungrily over the water taking some unknown nourishment⁸⁶ from the water. The few samples that have been caught have had in their stomachs only a slimy⁸⁷ substance from the sea and bits of pumice⁸⁸ from some far off volcanic island or reef. They lead a solitary⁸⁹ life and will approach a ship only at night when they are attracted by the lights. These tiny birds flash through the aura⁹⁰ of a mainmast light, circle it for several minutes and then vanish back into the darkness. There is something ascetic⁹¹, shy, courageous and aloof about the storm petrel.

The man-o'-war hawk has a huge wingspread of over seven feet and tiny atrophied⁹² feet which are useless for anything but perching⁹³ which it has brought to a fine art. It can perch on the smallest twig and sleep peacefully. It can live almost indefinitely⁹⁴ in the air, its great wings incredibly⁹⁵ sensitive to any thermal rise or fall, hoarding its energy as it floats without the slightest exertion⁹⁶. I once watched as a man-o'-war hawk spied⁹⁷ a surface fish and stalked⁹⁸ it. It was a miracle of tactical⁹⁹ ingenuity. It dropped effortlessly and soundlessly from a few hundred feet, carefully keeping its shadow from falling over the fish. If the man-o'-war land on level ground they will not be able to take off. But even the slightest incline will allow them to raise their vast¹⁰⁰ wings, as the morning sun creates fugitive¹⁰¹ and tiny thermal¹⁰² layers, and begin to soar. They land in a gentle silent descent¹⁰³ which is actually a life-and-death calculation and maneuvering¹⁰⁴ for advantageous spots.

86. 1) energy 2) diet 3) food 4) refreshment
87. 1) slippery 2) miry 3) greasy 4) sloppy
88. 1) powder 2) sandstone 3) volcanic-rock 4) pummel
89. 1) single 2) unhappy 3) abandoned 4) excluded
90. 1) beam 2) aroma 3) glow 4) pinnacle
91. 1) abstaining 2) intemperate 3) odd 4) sterile
92. 1) dead 2) starved 3) atrocious 4) shrunk
93. 1) sitting 2) climbing 3) standing 4) hanging
94. 1) securely 2) precisely 3) indifferently 4) indeterminately
95. 1) uncertainly 2) plausibly 3) unbelievably 4) questionably
96. 1) effort 2) struggle 3) gain 4) exhaustion
97. 1) glimpsed 2) searched 3) hid 4) led
98. 1) deserted 2) observed 3) waited 4) approached
99. 1) polite 2) concise 3) diplomatic 4) planned
100. 1) infinite 2) great 3) indefinite 4) abundant
101. 1) deceptive 2) elusive 3) permanent 4) trapped
102. 1) cool 2) vacuum 3) heat 4) temperature
103. 1) slump 2) slide 3) drop 4) downcast
104. 1) attacking 2) scheming 3) defending 4) climbing

GO RIGHT ON TO NEXT PAGE

Although the battle for survival is incessant among the
 105
 wandering birds, some manage to survive and the species continues.
 106
 But those birds which have given up the fight and retreated to the
 107
 safety of lonely Pacific islands have had strange and bitter things
 happen to them. The Dodo suffered the worst of nature's penalties:
 it utterly vanished. Other birds, stupefied by generations of
 108
 undangerous and languid life, lost their fear of men and were quickly
 109
 caught and eaten. The white man also brought the black rat against
 which the dumb and slow birds had no defense. The rats glutted
 110
 on their compliant victims and in some cases would decimate a whole
 111
 clump of trees of its bird population in a single day.

Because of its fish diet the guano of sea birds is an excel-
 112
 lent fertilizer. The volume of guano deposited by the billions of
 113
 birds is staggering. Fortunately most of it is washed out to sea
 114
 by heavy rains. If the guano were allowed to accumulate undis-
 115
 turbed some islands would rise thousands of feet into the air and
 be natural resources of the first magnitude. On some islands the
 116
 droppings of birds have been mixed with the broken-down coral sand
 117
 and been converted into a high grade of lime phosphate. Islands
 with such deposits are methodically mined, great slashes cut into
 118
 them by bulldozers and dump trucks. Such islands provide the only
 119
 truly hideous landscapes in the Pacific.
 120

105. 1) continuous 2) frequent 3) random 4) immoral
106. 1) birds 2) breed 3) sample 4) items
107. 1) celebrated 2) withdrawn 3) recessed 4) proceeded
108. 1) amazed 2) sensitized 3) shocked 4) dulled
109. 1) fluent 2) alert 3) listless 4) energetic
110. 1) emptied 2) stuffed 3) gulped 4) devoured
-
111. 1) dismantle 2) decipher 3) devastate 4) demise
112. 1) nest 2) excrement 3) food 4) carcass
113. 1) deposed 2) entrusted 3) saved 4) dropped
114. 1) overwhelming 2) overloaded 3) hesitating 4) dangling
115. 1) receive 2) preserve 3) adjust 4) amass
-
116. 1) revenues 2) means 3) assets 4) credits
117. 1) pattern 2) power 3) dimension 4) height
118. 1) neatly 2) systematically 3) symmetrically 4) routinely
119. 1) caves 2) fractures 3) divisions 4) gashes
120. 1) ugly 2) mechanical 3) pungent 4) secret

STOP HERE, UNTIL TOLD TO GO ON. IF YOU HAVE EXTRA TIME, YOU MAY GO BACK OVER THIS SECTION.

Appendix B

Table A
Descriptive Statistics of All Tests Used in
This Study

<u>Test</u>	<u>Number of Subjects</u>	<u>Mean</u>	<u>Standard Deviation</u>
<u>Michigan Adult Reading Test</u>			
- Mark (M)	128	281.3	59.1
- Comprehension Check			
- stdC (forms B+C)			
- with reading	65	54.7	10.2
- without reading	30	46.4	12.6
- modC (forms A+D)			
- with reading	63	58.0	15.4
- without reading	30	46.0	10.0
- Power Vocabulary (Vp)	128	75.2	11.1
- Context Vocabulary (Vc)			
- with blanks	98	73.5	14.2
- with words	30	77.5	14.9
<u>Diagnostic Reading Test</u>			
- Rate (R)	126	307.4	66.8
- Total Comprehension (TC)	128	79.2	11.9
- Vocabulary (V)	128	89.6	5.8
Understanding Communication (UC)	128	90.5	7.9

Appendix B

Table B
Reliability Estimates for all Tests Used in
This Study

Test	Split-half (+ Spearman- Brown)	Kuder- Richard- son #21	From Test Manual
<u>Michigan Adult Reading Test</u>			
- stdC (forms B+C)	.523	.760	
- modC (forms A+D)	.757	.901	
- Power Vocabulary (Vp)	.735	.854	
- Context Vocabulary (Vc)			
- with blanks	.849	.910	
- with words	.920	.933	
<u>Diagnostic Reading Test</u>			
- Total Comprehension (TC)		.894	.76-.85
- Vocabulary (V)		.748	.81-.91
<u>Understanding Communication</u>		.795	.91
stdC - all items		.760	
- content word items		.384	
- function word items		.275	

Appendix B

Table C
Correlation Matrix of All Tests Used in This Study

C	Vp	Vc	R	TC	V	UC
M stdC .259 modC .216	.391	A,B,C .352 D .119	.676	.100	-.001	.271
C	stdC .356 modC .492	A .508 B+C .432 D .592	stdC .469 modC .109	stdC .511 modC .670	stdC .571 modC .527	stdC .612 modC .520
Vp		A,B,C .412 D .792	.173	.263	.352	.466
Vc			A,B,C .432 D -.076	A,B,C .366 D .590	A,B,C .349 D .404	A,B,C .604 D .619
R				.112	.088	.306
TC					.370	.668
V						.510

Correlations needed to be			Key -	
Group	N	sig. .05	sig. .01	
A	33	.349	.449	M = Mark (rate estimate) MART
D	30	.361	.463	C = Comprehension Check MART
stdC	65	.250	.325	Vp = Power Vocabulary MART
modC	63	.250	.325	Vc = Context Vocabulary MART
A,B,C	98	.205	.267	R = Rate (rate estimate) DRT
All	128	.174	.228	TC = Total Comprehension DRT
				V = Vocabulary DRT
				UC = Understanding Communication
				DRT = Diagnostic Reading Test
				MART = Michigan Adult Reading Test

Appendix B

Table D
Correlation Matrix of Content and Function Items

Test	stdC (N = 65)		stdC (N = 63)	
	Content (n = 52)	Function (n = 48)	Total (n = 100)	(n = 50)
DRT - TC	.353 ¹	.290	.511	.670
DRT - Story Comprehension	.350	.186	.475	.650
DRT - Paragraph Comprehension	.274	.166	.403	.425
UC	.465	.284	.612	.520
MART - Vp	.318	.256	.492	
MART - Vc	.363	.268	.432	

¹For all coefficients this table $r = .250$, $p < .05$; $r = .325$, $p < .01$
Correlation between content and function items of stdC; $r = -.233$

Bibliography

- American Psychological Association Publication Manual, Washington, D. C., 1957 (rev.).
- Bormuth, J. R. Cloze tests as measures of readability and comprehension ability. Unpublished doctoral dissertation, Indiana Univer., 1962.
- Bormuth, J. R. Experimental applications of cloze tests. Paper read at Intl. Read. Assoc., Philadelphia, May 1964.
- Burdick, E. The Blue of Capricorn, Boston, Houghton-Mifflin, 1961.
- Coleman, E. B. & Blumenfeld, J. P. Cloze scores of nominalisations and their grammatical transformations using active verbs. Psychol. Rep., 1963, 13, 176-179.
- Fan, C-T. Item analysis table. Princeton: Educational Testing Service, 1952.
- Fletcher, J. E. A study of the relationships between ability to use context as an aid in reading and other verbal abilities. Unpublished doctoral dissertation, Univer. of Washington, 1959.
- Jenkinson, Marion D. Selected processes and difficulties of reading comprehension. Unpublished doctoral dissertation, Univer. of Chicago, 1957.
- Manis, M. & Dawes, R. M. Cloze scores as a function of attitude. Psychol. Rep., 1961, 9, 79-84.
- Rankin, E. F. An evaluation of the cloze procedure as a technique for measuring reading comprehension. Unpublished doctoral dissertation, Univer. of Michigan, 1958.
- Rankin, E. F. The cloze procedure - its validity and utility. In Causey, O. S. & Eller, W. (Eds.), Starting and improving college reading programs. Yearb. Natl. Read. Conf., 1959, 8th, Pp 131-144.
- Rankin, E. F. The relationship between reading rate and comprehension. In Bleismer, E. P. & Staiger, R. C. (Eds.), Problems, programs and projects in college and adult reading. Yearb. Natl. Read. Conf., 1961, 11th, Pp 1-5.
- Rankin, E. F. The definition of reading comprehension. In Raygor, A. (Ed.), College and adult reading. Yearb. North Central Read. Assoc., 1962, 1st, Pp 15-31.

Bibliography (continued)

- Ruddell, R. B. An investigation of the effect of the similarity of oral and written patterns of language structure on reading comprehension. Unpublished doctoral dissertation, Indiana Univer., 1963.
- Smith, D. E. P. & Carrigan, Patricia. Manual for the SA-S senior and junior scales. Univer. of Michigan, 1957.
- Taylor, W. L. "Cloze procedure": a new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.
- Taylor, W. L. Application of "cloze" and entropy measures to the study of contextual constraint in samples of continuous prose. Unpublished doctoral dissertation, Univer. of Illinois, 1954.
- Taylor, W. L. "Cloze" readability scores as indices of individual differences in comprehension and aptitude. J. appl. Psychol., 1957, 41, 19-26.
- Thurstone, Thelma. Understanding Communication; verbal comprehension. Univer. of Chicago, Industrial Relations Center, Chicago, 1959.
- Triggs, Francis (Ed.). Diagnostic Reading Tests, Mountain Home, N. C., 1948.
- Weaver, W. W. An examination of some differences in oral and written language using the cloze procedure. Unpublished doctoral dissertation, Univer. of Georgia, 1961.