CLOZE FORMAT FOR FOURTH YEAR PUPILS

by

DONALD PERCY WALMSLEY

B.Ed., University of Victoria, 1968

A THESIS SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Faculty

of

Education

ACCEPTED

FACULTY OF GRADUATE STUDIES

[redacted]

Assoc. DEAN

DATE 19 August 1974

We accept this thesis as conforming to the required standard

[redacted]

UNIVERSITY OF VICTORIA

August 1974

Supervisor: Dr. R. D. Armstrong

ABSTRACT

Matching the readability level of available print materials with the reading ability of pupils was identified as a major problem of classroom teachers. Cloze procedure was found to be one method of making a suitable match. The advantages of cloze procedure were: (a) it was quick to prepare, administer, and mark and interpret; and (b) it measured the readability of the material and the reading ability of the subject with the same instrument. Cloze test scores could be used to indicate whether or not the material tested was within the subject's reading capability.

The purpose of the present study was to develop a simple, reliable cloze test format which had concurrent validity as a reading test.

Three aspects of cloze format were tested: (a) deletion rate, (b) test length, and (c) number of cloze tests required to adequately test a selection.

Data were used from a total of 132 subjects at the fourth year level in the two phases of the study. Three groups of 34 pupils took part in the first phase. Each wrote 1 set of 2 cloze test forms 300 words long. Each set used a different deletion rate. Reliabilities of the sets were compared. Next, reliabilities for lengths ranging from

30 to 300 words were compared. One group of 30 pupils took part in the second phase of the study. This group wrote three cloze test forms based on one selection and a standardized reading test. Their teacher ranked them according to her perception of their reading abilities. Reliabilities over 1, 2, and 3 forms were compared. As a result of the three comparisons cloze format (CF) consisting of one test form 250 words long with a 1-in-5 deletion rate was determined.

Split-half and alternate form reliabilities of CF were tested for significance and compared with similar reliabilities of a standardized reading test. The standardized reading test scores and the teacher's ranking were considered to be measures of reading ability. Correlations between these and CF were tested for significance.

Split-half reliability was .897, and alternate form reliability varied from .958 to .965. Correlations between CF and the standardized reading test varied from .945 to .952. The Spearman Rho correlation between CF and the teacher's ranking was .925. All results were significant at the .01 level of confidence.

It was concluded that CF was simple and reliable and that it had satisfactory concurrent validity as a test of reading ability.

It was noted that, to be useful for pupil placement in

suitable reading materials, the individual CF scores had to be translated into readability scores. A table, compiled from the literature and not an aspect of this study, was provided for this purpose.

Examiners: [redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Dr. R. D. Armstrong, Chairperson of the Supervisory Committee, for his guidance, support, and extensive patience during the preparation of this thesis. The encouragement of Mr. C. Campbell and the advice of Dr. T. D. Johnson are gratefully acknowledged. Special thanks are extended to Dr. D. Chabassol whose assistance was largely responsible for the task having been begun.

The author also wishes to acknowledge the help of Dr. W. Muir and Dr. H. Taylor with the statistical design.

Recognition is given to the pupils, teachers, and principals of the Hope School District who participated in this study.

Finally, the author is indebted to Dr. G. K. Pederson, Dean of Education, and to the Faculty of Graduate Studies for the time extension which made the writing of this thesis possible.

To my family,

for their understanding

CHAPTER I

INTRODUCTION TO THE PROBLEM

Introduction

Many educators believe that

> The broad prime aim of the public school system should be to foster the growth and development of every individual to the end that he/she will become and be a self-reliant, self-disciplined, participating member with a sense of social responsibility within a democratic society [British Columbia Teachers' Federation, 1974, p. 56].

The development of language skills in individuals is considered basic to the attainment of this aim. Facility must be developed in both oral and written language--listening, speaking, reading, and writing--before an individual is able to evaluate critically and react rationally to the messages he/she receives. The development of these language skills is considered so important that our school system allocates approximately one-half of its instructional time to the teaching of language arts. Even at the secondary level, if some of the time spent in teaching social studies, science and mathematics is considered as being indirectly associated with the teaching of language arts, the half time figure holds. However, it can be readily

determined by checking language arts test results at any grade level that a significant number of our students do not achieve a satisfactory mastery of these skills. Some students, in fact, seem never to reach a level of development which would allow for the attainment of the broad prime aim set out at the beginning of this section.

The lack of success of certain students has been the major spur to the search for new teaching methods which might help low achieving students increase their facility in language skills.

The present study concerned itself with the development of a cloze test format which could be used to improve reading instruction by providing a readily accessible measure of readability of print material and reading ability of the pupil.

The Problem

The goal of the elementary school reading program is to develop each pupil's reading capability. The pupil provided with reading material which is too easy wastes his/her time, whereas the pupil forced to attempt to read material which is too difficult becomes frustrated and too often loses confidence in his/her ability to read. The major aim of every reading instructor is, therefore, to find a pupil's reading ability and to provide him/her with material at that level. If the diagnosis and prescription are accurate, the

child will experience continuous and successful learning (Bormuth, 1967a; Zintz, 1970). The problem, of course, lies in finding ways to make accurate diagnoses and prescriptions.

Diagnosis

The first aspect of this problem, the diagnosis, relates to the child. The diagnosis must measure the level of his/her reading ability. Reading ability is the measure of the child's level of development of word recognition skills, vocabulary development and comprehension. Raw scores representing these measures can often be translated into grade equivalent scores. Such scores provide an instructor with a good indication of the level of difficulty of the materials the child should be reading.

In this section, diagnosis is used in a general sense. It does not refer to the type of diagnosis which would take place in a reading clinic.

Prescription

The second aspect of the problem, the prescription, relates to the material. Clearly, some books are more difficult to read than others. Various methods can be used to measure the relative reading difficulty or readability of written materials. Such methods most often utilize sentence lengths and the relative frequency of the words used as indices of difficulty. Raw scores representing such measures can also be transformed into grade equivalent scores.

Once a teacher has a measure of a child's reading

ability along with a series of books for which he/she has readability coefficients, a match of the child's reading needs with suitable materials can be made.

The Diagnostic-Prescriptive Methods of Pupil Placement

Reading teachers have used a range of methods to determine the match described above. These methods have varied from the highly informal, such as intuition, to the very formal use of standardized tests.

Some teachers claim that, as a result of close contact with a child, they can place him/her accurately without resorting to the use of tests. Without arguing the validity of this claim, it seems to presuppose that there is a very small number of pupils per class and that the teacher has a knowledge of the readability levels of a large number of print materials.

Another method of determining pupil placement is the informal reading inventory (Zintz, 1970). This method requires the teacher to administer individual word recognition and comprehension tests to his/her pupils. Such a method provides a teacher with objective data for determining reading ability. This knowledge, linked with an understanding of the readability of various materials, should lead to the use of suitable materials. However, since approximately one-half an hour per pupil is required to administer the test, extensive use of the method is prohibitive.

The more formal methods of determining reading placement

generally use a standardized testing procedure to diagnose reading ability along with a readability formula to prescribe suitable reading materials. Standardized tests have the advantage of taking less time to administer than do the informal methods.

However, such tests are said to be misleading in that they often misrepresent the abilities of individual students within the class (Jones & Pikulski, 1974). These tests also take a great deal of teacher-time to mark and interpret. Further teacher-time is then required to find reading materials at suitable readability levels. This task is made more difficult, first, because so many books do not indicate readability levels, second, because many readability formulae are very time consuming to apply and, third, because the application of one readability scale may yield quite a different readability coefficient than a second. Klare (1963), for example, identified 31 different formulae and others have since been published.

An exception to the heavy time commitment required to use the standardized test-readability formula placement method is the Science Research Associates' Reading Laboratory Kit approach (Science Research Associates, 1969). An SRA kit measures pupil reading ability and determines placement within the kit in virtually one step.

The close connection of the reading ability measure to the readability scale found in the SRA kit illustrates the

single greatest advantage of cloze procedure.

Cloze Procedure

A second formal method of determining pupil placement is through the use of cloze procedure. This method has the advantage of measuring both reading ability and readability with the same test. In addition, cloze tests are simple to prepare, administer and mark.

Cloze procedure consists of (a) taking an oral or written passage, (b) deleting a number of words from the passage in some predetermined manner, (c) replacing the deleted words with blanks, (d) giving the mutilated passage to a subject, and (e) asking the subject to fill in the blank spaces with the words he/she thinks were deleted. A cloze score is the total number of correct responses made by the subject over the passage. The sentence, "A rider on a _____ horse left us with _____ strange message." is an example of a short passage to which the cloze procedure has been applied.

The procedure got its name from Taylor (1953) who compared the "clozing" of the blanks with the gestalt psychology concept of "closure." Closure is explained as the human tendency to complete mentally a partially drawn figure. For example, subjects tend to see a partially drawn circle as a circle. The circle is completed mentally.

The ability to cloze a mutilated language passage seems to result from the fact that language meaning is conveyed at

three different levels: syntactical (structural or grammatical), lexical (dictionary or denotative), and social-cultural (experiential or connotative) (Anderson, 1971). If one level is incomplete as a result of the cloze mutilation, another level can help to supply the missing information. In the sentence, "Bob found _____ baseball glove but his _____ was broken." various clues help to cloze the gaps. A knowledge of English structure, developed through usage, allows a subject to rule out all word classes except pronouns and articles as replacements for the first blank. Either "a" or "the" would fit, however, the use of "but" and "his" later in the sentence probably indicate a parallel structure earlier in the sentence. Within the remaining group, pronouns, plurals can probably be eliminated because there is only one glove. Finally, the fact that "Bob" is masculine probably indicates that if the missing word is a pronoun, it will be "his." To fill the second blank, again a knowledge of English indicates that only one word class, nouns, will fit. The word must be singular. If the subject had some experience with baseball, a clue from earlier in the sentence, probably he/she would associate "bat" with "broken." Consequently, an understanding of language, which results from the ability to utilize the three levels mentioned earlier, can be measured by a cloze test score. Various studies have been made to relate this reading ability score to the readability level of the

material tested. It is generally accepted (Mork, 1971) that an individual's cloze score of 45% or more indicates that the material is within his/her reading ability. In other words, to determine whether or not a particular selection is within the reading ability of a class (or an individual), a teacher needs only to construct a close test on a passage taken from the selection, administer it to the class, mark it, convert the cloze scores into percentage scores, and determine by inspection how these scores relate to the 45% criterion. If a large number of scores are below 45%, the material is too difficult, but if the scores tend to equal or exceed 45%, the selection is readable. Cloze score percentages of 60 or more are considered to indicate that the material is too easy for the group tested.

The advantages of cloze procedure make it an ideal device not only for placement but also for ongoing evaluation.

This last feature is of great importance because to obtain maximum success, a reading program must not only determine the initial placement of a pupil but must continually re-evaluate the pupil's progress to insure that optimum learning is maintained. In the author's opinion, such follow-up is easier using the cloze procedure than it is using one of the other methods described.

Purpose of the Study

On the assumption that pupils must be given suitable materials if they are to reach their reading potentials, the author tested various cloze formats to find a method of matching pupil reading ability with material readability.

The purpose of this study was to test three aspects of cloze test format in order to recommend a simple, reliable and valid procedure by which classroom teachers at the fourth year level might select suitable reading materials.

The three aspects tested were deletion-rates, cloze passage lengths, and number of cloze passages necessary for the testing of a given selection. Deletion rates tested were one-in-five, one-in-eight, and one-in-eleven. Cloze passage lengths tested were 30 words, 50 words, 100 words, 150 words, 200 words, 250 words and 300 words. The number of passages per selection tested were 1, 2, and 3.

The author believed that shorter cloze tests would be more attractive to classroom teachers than longer tests since less time would be required to prepare, administer, and mark them. On the other hand, the literature suggested that greater length and frequency of tests increased the reliability, and therefore the possibility of validity, of cloze procedure (Taylor, 1953; Bormuth, 1965). The most critical aspect of this study was to find a cloze test format which could be defended in terms of both its simplicity and its reliability.

Definition of Terms

Cloze procedure. Cloze procedure was the method of deleting a predetermined number of words from a language passage, asking a subject to supply the missing words, and measuring his/her success.

Readability. The measure of the reading difficulty of a prose selection was called readability.

Reading ability. The measure of a pupil's level of development of word attack, vocabulary, and comprehension skills was considered to be his/her reading ability.

Deletion rate. A deletion rate of one-in-five meant that every fifth word was deleted and replaced by a blank of standard length. Similarly, one-in-eight and one-in-eleven deletion rates referred to the deletions of every eighth and eleventh words and their replacement by blanks of standard length.

Blanks of standard length. In the cloze tests constructed, each blank was the same length to avoid giving clues as to the length of the word deleted. In the present study all blanks were 15 spaces long to provide the younger pupils adequate room for their answers.

Selection. A selection was a short story, novel, or essay.

Passage. A passage was a part of a selection. In this study passages varied in length from 30 to 300 words.

Basal reader. A basal reader was considered to be a

collection of passages selected on the basis of a readability formula for use at a specified grade level.

Teacher's judgement. The teacher's judgement was the rank order in which the teacher placed her pupils according to her perception of their reading abilities.

Hypotheses

The purpose of the present study was to develop a simple cloze test format. It was recognized that to have any value, such a test format would have to be both reliable and valid. This led to the development of the hypotheses.

Hypothesis I

The cloze test format developed for fourth year pupils has split-half and alternate form reliability.

Hypothesis II

The cloze test format developed has concurrent validity as a test of reading ability.

Cloze Test Format Development

Before testing these hypotheses, the present study had to develop a simple cloze test format. To do this, three questions had to be answered.

1. Of the deletion rates, one-in-five, one-in-eight, and one-in-eleven, which was the maximum rate which still possessed satisfactory reliability?

2. Of the cloze test lengths of 30 words, 50 words, 100 words, 150 words, 200 words, 250 words, and 300 words, which

was the shortest length that still possessed satisfactory reliability?

3. Of the numbers of cloze tests per passage, 1, 2, and 3, which was the lowest number which still possessed satisfactory reliability?

The present study, therefore, had two aspects. First, it was to develop a simple cloze test format and, second, it was to test the format.

Design of the Study

Instruments

For this study, the author used nine cloze tests and "Form 2M, Survey D," of the Gates-MacGinitie Reading Tests (Gates & MacGinitie, 1965). The cloze tests were based on passages from two selections from a fourth year basal reader. Six tests were developed from the first selection to answer the questions relating to deletion rate and length. These tests were divided into three sets having parallel forms. Each set used a different deletion rate. The other three cloze tests were based on three passages from the second selection. All three were the same length and used the same deletion rate.

Procedure

Four groups of fourth year pupils from the Hope School District participated in this study. In the first phase, data from 34 pupils in each of the first three groups were

used. In the second phase, data from 30 pupils in the last group were used.

The Sample

All of the fourth year pupils from two elementary schools in Hope were selected for the first phase of the study. The 82 pupils from the first school were assigned to two groups. The 35 pupils from the second school were assigned to the third group. Each group of pupils wrote one set of two cloze test forms. From each of the three sets of paired cloze tests thus produced, the author selected 34 pairs.

Table I shows how the cloze tests and subjects were arranged in Phase I.

In the second phase, all 33 fourth year pupils from a third school wrote three other cloze test forms as well as the Gates-MacGinitie Reading Tests, "Form 2M, Survey D " (Gates & MacGinitie, 1965). This group was also ranked by its teacher. At the end of Phase II, the author had three sets of measures of each individual in group four: cloze test scores, Gates-MacGinitie scores, and teacher's judgement. Table II indicates the subjects and measures from the second phase.

Administration

All pupils were given two sample tests to familiarize them with the cloze task. All formal tests were administered on successive days immediately following recess. Phase I

TABLE I

Sets of Cloze Test Forms and Subjects: Phase I

Test Forms				Subjects			
Set	Cloze Forms	Deletion Rate	Passage	School	Group	Subjects	Paired Cloze Forms
One	A	1-in-5	1	1	1	41	34: 7 pairs were
	D	1-in-5	2	1	1	41	randomly removed
Two	B	1-in-8	1	1	2	41	34: 7 pairs were
	E	1-in-8	2	1	2	41	randomly removed
Three	C	1-in-11	1	2	3	35	34: one S wrote
	F	1-in-11	2	2	3	35	only one form

TABLE II

Subjects and Measures: Phase II

Subjects	Measures
Group 4, School 3 30 subjects: 33 began but 3 did not complete all tests	Cloze tests G, H, I Gates-MacGinitie Teacher's judgement

took two days and Phase II took five days. Subjects were given time to complete the cloze tests but the Gates-MacGinitie tests were administered in accordance with the "Teacher's Manual" (Gates & MacGinitie, 1965).

All tests, except the sample tests, were returned to the author, unmarked, as soon as was practical after the testing had taken place.

Analysis of the Data

After the tests in Phase I had been marked and checked, correlations were calculated for the paired test forms in each of the three sets. These correlations were tested for individual significance (Popham, 1967) and for significant differences (Ferguson, 1966). Based on statistical and practical considerations, a deletion rate was selected for the cloze test format.

Coefficients were then calculated for the correlations of the paired forms in one set over lengths ranging from 30 to 300 words. All correlations were tested for individual significance and for significant differences. The cloze test length was then selected for the format.

In the second phase, split-half correlations were calculated on Form G, Forms G plus H, and Forms G plus H plus I. These figures were tested for individual significance once they had been corrected for test length. The number of cloze tests required to test a selection was then

chosen.

At this point in the study, the cloze test format had been selected based on both statistical and practical considerations.

The hypotheses were then considered. Correlations were calculated to test both split-half and alternate form reliability. These correlations were tested for significance and then compared with similar reliabilities reported by a recognized standardized test.

Concurrent validity was tested by correlating cloze test and Gates-MacGinitie test scores and then testing this coefficient for significance. Cloze scores were then correlated with the teacher's judgement rank-ordering. Similarly, the Gates-MacGinitie scores were correlated with the teacher's judgement. These correlations were tested for significance and compared.

In all but the last two cases, correlations were computed using the Pearson product-moment correlation formula (Popham, 1967, p. 88). The last two correlations dealt with rank-ordering so they were computed using the Spearman Rho formula (Popham, 1967, p. 314).

Limitations of the Study

1. The major limitation of this study was the lack of sensitivity of the correlation coefficients and of the procedures for comparing correlation coefficients. That is,

while a correlation coefficient of .84 was higher than one of .67, applications of statistical treatments often did not show that the difference between them was statistically significant.

2. Another limitation was associated with this. The models for testing significant differences between correlation coefficients, as outlined in Ferguson (1966), did not quite fit some of the situations encountered in the present study. In order to have some method of testing for significant differences between correlations these models were twice used in situations somewhat different than those outlined by Ferguson.

3. The second phase of this study, including the teacher's judgement, took place in November. It is quite possible that a ranking later in the year, after the teacher had an opportunity to learn more about her pupils, might have been different.

4. The small size of the sample, especially in the second phase, raises some doubt about the degree to which the results should be generalized.

5. In spite of attempting to base all format choices on statistical data, this was often not possible. As a result, some choices were arbitrary. It is possible that had different choices been made, they would have resulted in similar findings for a different format.

6. The present study explored only some of the alterna-

tives in each of the areas of deletion rate, test length, and number of cloze tests required to test a selection. Other alternatives might have resulted in choosing a different format.

Organization of the Study

This chapter has attempted to set out the framework of the present study. The remainder of the study is organized as follows:

Chapter II : Review of Related Literature

Chapter III: Design of the Experiment

Chapter IV : Analysis and Results

Chapter V : Summary, Conclusions and Implications.

CHAPTER II

REVIEW OF RELATED LITERATURE

The first part of this chapter deals with the literature on readability, readability measures, reading ability, and measures of reading ability. The second part contains a discussion of cloze procedure including its use as a measure of reading ability and readability.

PART I

Readability and Reading Ability

Readability is the reading difficulty of any written selection. It is clear that some selections are far more difficult to read than others. It is equally clear that a measure which made possible the ranking of various selections according to their readability would be a valuable aid in helping to decide which reading materials were suitable for students at particular levels of reading development.

The need for such measures is paramount. Bormuth (1967b) stated

> No one knows how much time is wasted each year on instructional materials that are too difficult for many students to understand. When measured in money, the waste may run into millions of

> dollars. But when measured in human terms, the waste becomes appalling. The costs of stunted education of children, the frustrations and negative attitudes arising from their failures to learn, and the contributions to the drop out rate with its consequences of unemployment and poverty result in a staggering drain on [a] nation's resources [p. 291].

Mork (1971) stated that the problem faced by a reading teacher lay in providing students materials that they would find challenging yet in which they would have a high probability of success. Jones and Pikulski (1974) summarized current educational thought in this area when they stated

> It is generally agreed that children benefit most from reading instruction which occurs in materials which are challenging and interesting and yet not so difficult as to be frustrating [p. 432].

The problem of providing pupils with suitable reading materials has been recognized for a very long time. One of the most important reasons for the creation of grade levels within schools was to meet this problem in a gross manner. With the re-introduction of ungraded schools through the adoption of the principle of continuous progress, the problem again seemed, to the author at least, to become more manifest. As a result, the levels system was devised in an attempt to provide students with suitable reading materials.

Along with the attempts to organize the school system so it would become more sensitive to the needs of its students, educators attempted to find devices which would indicate suitable reading materials for students. Earliest

methods attempted to use a student's score on a standard reading test to predict suitable material as determined by a readability formula.

The development of the informal reading inventory in which a student read graded selections to a teacher who then determined which material was suitable was a further attempt to ensure that student and reading materials matched. The most recent device used to determine readability and pupil placement has been cloze procedure.

Readability Measures

Readability Formulae

A readability formula is used to determine the degree of reading difficulty of a passage written in English. A formula identifies various factors inherent in English, counts their frequency and converts the score thus found into a readability coefficient. Most formulae then convert these readability coefficients into grade equivalent scores. This is generally done by having a sample population write a series of tests, first on material ranked by a readability formula and, second, on standardized reading tests. By comparing the results of each, grade equivalents are established. Some formulae omit this step and instead compare their scores with scores established by earlier formulae.

The usual factors which have been considered by readability formulae are vocabulary load, sentence structure,

affixes, idea density and human interest (Lorge, 1949).

Following is an explanation of how a typical formula is applied. Vocabulary load is determined by the diversity and rarity of the words used. To determine this, the words in the passage are usually checked against a list of commonly occurring words such as the Dolch List (Deschant, 1970). The greater the number of common words contained in a passage, the easier it is judged to be. Average sentence length over the passage is then calculated. The shorter the average sentence length, the easier the passage. Similarly, the number of ideas in a passage are counted with the assumption that fewer ideas mean easier material. The number of personal pronouns are then counted to determine human interest. In this case, the more such pronouns counted, the easier the passage is considered to be.

Klare (1963) identified 31 readability formulae. Not all counted precisely the same factors as those described above but it is clear that there is simply not sufficient space to discuss them all here.

Readability formulae have constantly been improved (Klare, 1966) but refinements made have been possible chiefly through the use of computers. However, to the classroom teacher who now has a proliferation of materials available, such complex formulae are usually not practical. A formula which is much more simple is desirable. Such a formula is, unfortunately, subject to a greater possibility

of error. One such formula was developed by Fry (Deschant, 1970, pp. 279-280). This formula counts only sentence and word lengths. Figure 1 shows this formula.

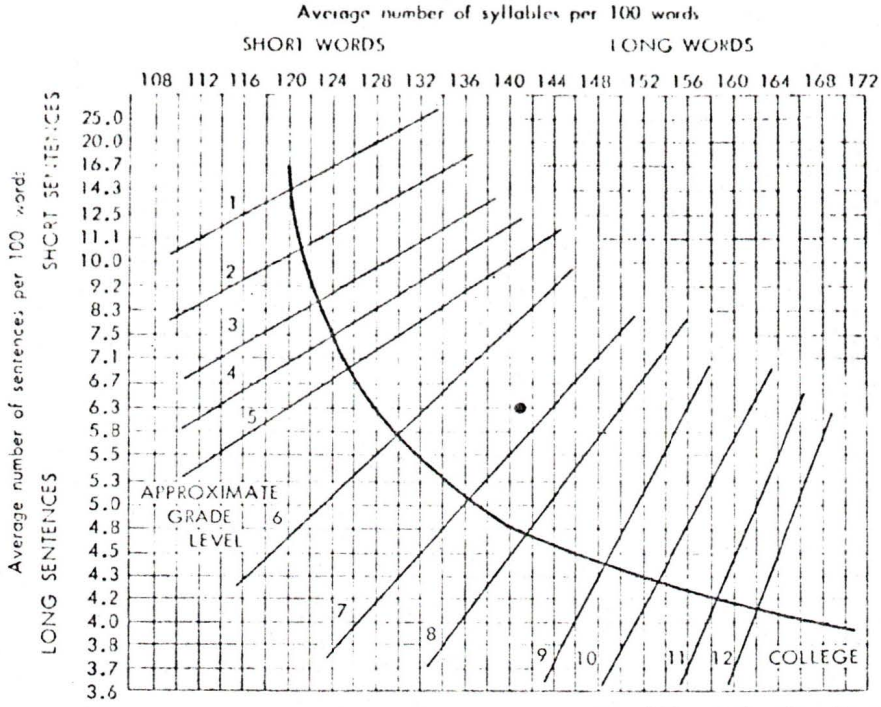
The merits of readability formulae are discussed later in this chapter.

Providing Students with Suitable Reading Materials

One method of determining the materials suitable for a group of students is to administer a standardized reading test to them and select for them material which is rated at the same readability level as their test scores indicate (Deschant, 1970). While this method has been used with individuals as well as groups it has tended to be impersonal and subject to error. Moskowitz (1971) stated that standardized tests were only a sample of a child's general performance under a controlled condition. As such, often very little information of value to the classroom teacher came from them.

Those who objected to the formality of standardized testing, the results obtained from it, and the type of instruction it tended to encourage, developed the informal reading inventory approach to determine pupil placement (Zintz, 1970).

A typical informal reading inventory contains a collection of graded reading materials which a student reads orally to his/her teacher in order that the teacher can determine the student's word recognition ability. A second



Graph for Estimating Readability

Directions: Randomly select three 100-word passages from a book or an article. Plot average number of syllables and average number of words per sentence on the graph to determine the grade level of the material. Choose more passages per book if great variability is observed and conclude that the book has uneven readability. Few books will fall in the gray area, but when they do, grade level scores are invalid.

Example:	*syllables*	*sentences*
first 100 words	124	6.6
second 100 words	141	5.5
third 100 words	158	6.8
average	141	6.3

Readability: 7th grade (see dot plotted on graph)

Figure 1 Fry readability formula (Deschant, 1970)

collection of graded materials is then read silently by the student. Following each selection the teacher asks a series of questions to discover the student's comprehension level. Typically, the graded materials are chosen from basal readers which have been rated by one or more readability formulae. On the basis of the student's word recognition and comprehension abilities, instructional materials are chosen (Veatch, 1959; Zintz, 1970).

Informal reading inventory scores usually indicate one of three reading development levels: independent, instructional, or frustration. The correct level is determined by converting the word recognition and comprehension scores to percentages and by comparing these percentages to those listed in Table III.

For example, a student who read 100 words orally making three mistakes and who read a selection silently and answered correctly eight out of ten comprehension questions based on it, would be said to be working with material at his/her instructional level (Zintz, 1970). Once a student's reading level is found, he/she is guided to suitable reading materials.

The most recently used measure for determining readability and pupil placement, and one which is quite different from the two described above, is cloze procedure. Part II of this chapter deals with it.

TABLE III

Reading Development Level as Determined from Word Recognition and Comprehension Test Scores Expressed in Percentages

Reading Scores	Reading Development Level		
	Independent	Instructional	Frustration
Word recognition score (%)	99% to 100%	95% to 99%	less than 95%
Comprehension score (%)	90% to 100%	75% to 90%	less than 75%

PART II

In the second part of this chapter an attempt is made to justify the use of cloze procedure as a measure of readability and reading ability. An effort is made to trace the development of cloze procedure. The dynamics of cloze procedure are discussed. Cloze procedure is compared to other methods of determining readability. Studies relating to the reliability and validity of cloze tests are summarized. Finally, other uses of cloze procedure as a testing or instructional device are discussed.

Development of Cloze Procedure

The first use of cloze procedure is generally attributed to Taylor in 1953. However,

> . . . interest in letter and word deletion can be traced to Ebbinghaus in the 19th Century and the studies of intelligence by E. L. Thorndike in this century [Hafner, 1966, p. 415].

The author found it interesting to speculate that the first users of oral cloze procedure were the censors who replaced offensive language with bleeps.

Whatever its beginnings, certainly the experiments by Taylor led to the development of cloze procedure as it is today. Taylor's definitions of cloze units and cloze procedure have been noted. A cloze unit is

> Any single occurrence of a successful attempt to reproduce accurately a part deleted from a "message" by deciding from the context that

> remains, what the missing part should be [Taylor, 1953, p. 416].

Cloze procedure is

> A method of intercepting a message from a "transmitter" (writer or speaker), mutilating its language patterns by deleting parts, and so administering it to "receivers" (readers or listeners) that their attempts to make the patterns whole again potentially yield a considerable number of cloze units [Taylor, 1953, p. 416].

From its beginning in 1953 cloze procedure has developed into a widely used tool. Scores of researchers have used it in varying testing situations and many more have used it as a teaching device.

Cloze and Closure

The early explanations of the dynamics of cloze procedure centered around the gestalt concept of closure.

Taylor (1953) explained that he derived the name "cloze" from the gestalt psychology term "closure." He explained that in the same way as a person viewing an incomplete circle mentally completed it, so a reader viewing an incomplete sentence mentally completed it. The human tendency, claimed Taylor (1953), is to complete a familiar but not quite finished pattern. The degree to which a reader was familiar with the language used in a selection would therefore determine how well he/she would cloze that selection.

Louthan (1965) agreed that cloze procedure could be explained by the gestalt concept. Oller and Conrad (1971)

supported this by saying that the restoration of words deleted from a selection of prose in order for the passage to make sense was a special use of the ability to complete broken patterns. Weaver (1965) and Anderson (1971) did not agree that cloze procedure was an aspect of closure. Weaver (1965) argued that cloze responses were cognitive rather than perceptual and that the cloze task was subjective rather than objective. Anderson (1971) agreed that cloze procedure was subjective and that it did not make use of objective information in the same way that filling-in-the-blanks or sentence completion tests did. He went on to argue that cloze procedure measured comprehension not "closure."

The matter of whether or not cloze is related to closure appeared to depend largely on an individual's interpretation of the gestalt definition. It seemed more productive to pursue an explanation of how the cloze procedure worked based on an examination of the English language.

Cloze Procedure as a Measure of Redundancy

Manis and Dawes (1961) explained that English is a language of redundancies and that these redundancies generally have been thought to be safeguards against errors in communication. Because English is a language of redundancies, a message blocked at one level is not necessarily blocked at

all levels. Due to the redundancies inherent in our language, parts of sequences can be anticipated. A person reading or listening with understanding constantly makes tacit predictions of what will come next (Peisach, 1965).

Anderson (1971) explained that there are three levels of language to be understood: syntactical, lexical and social-cultural. Understanding is complete only when one understands all three levels but at the same time, understanding of one level aids understanding of another.

Through syntax, a sentence such as, "A man is coming." indicates the singular number three times: a, man, is (Taylor, 1953). A number of studies have pointed out the importance of syntax to a general understanding of our language. Moores (1970) found that deaf students showed much lower levels of syntactical understanding than did their counterparts who were able to hear. Even though lexical understanding appeared about equal in both groups, comprehension was higher in the group that was able to hear than it was in the group of deaf students. A study by Marshall (1970), which also involved deaf students, showed that the development of syntax and semantics were closely related. In a study done involving emotionally retarded students, it was shown that a lack of syntactical understanding impaired the development of lexical understanding (Semmel, Barritt, & Bennett, 1970). The author has noted, subjectively, similar occurrences with students for whom

English was a second language. Such students simply did not appear to have developed an ability to make use of all the non-lexical information available.

A lexical or dictionary understanding of English is well known to every student. In an effort to develop such an understanding in their students, teachers have used countless vocabulary drills and comprehension questions. A lexical understanding of "ducks" would likely lead one to cloze the sentence, "A duck has _____ feet." with the word "webbed." Previous experience with ducks, or an earlier part of the passage mentioning ducks, might have caused some other response such as "gray" or "flat." Such a response, however, could be attributed, in part at least, to social-cultural factors (Taylor, 1953).

Perhaps the following is a better example of the third level of understanding. The word used to cloze the sentence, "The road was paved with _____." would depend a great deal on the type of paving material commonly used in one's area or the fairy tales of one's youth. The importance of this level of understanding to a general understanding of language has been recognized by the developers of the "Culture-free" intelligence tests (Cattell & Cattell, 1959).

A further observation related to the cloze process was made by Weaver (1965, p. 115). Weaver said that cloze procedure demanded the person taking the test to alternate constantly between reader and writer. That is, at one point

the subject would be decoding (as he/she read the unmutilated section) but at another point the subject would be encoding and thus using thought processes common to writers.

Cloze procedure, then, is possible because English conveys meaning at various levels. A cloze score indicates how well the message of the sender was understood by the receiver.

Cloze Procedure and Readability

In support of the final statement of the last section, Hafner (1966) and Levine (1971) stated that the degree of success a reader has in completing a cloze test marks the degree of understanding he/she has of the passage. On the basis of this it follows that a cloze test score indicates whether the passage is above, below, or at the reader's ability level.

Cloze Procedure as a Readability Measure

Early users of the cloze procedure used cloze test scores to rank the readability of various selections and compared the results of the cloze rankings with rankings of the same selections by the applications of readability formulae of the Flesch and Dale-Chall type. In fact, the first use of cloze procedure did just that (Taylor, 1953). In what has become a fairly typical study, Taylor took three passages from Flesch's How To Test Readability, constructed cloze tests from them and computed the Dale-Chall scores on

them. The two readability formulae and the cloze test mean scores ranked the passages in exactly the same order. As recently as December, 1973, reports of similar experiments and findings have verified the propriety of cloze procedure as a readability measure (Galloway, 1973).

Cloze Procedure as a Placement Device

Having determined that cloze procedure could be used to rate the readability of various English selections, researchers attempted to answer the second part of the readability question: how could students be provided with suitable materials?

A study done by Bormuth (1967a) is typical of the work done in this area. Bormuth's goal was to find the cloze test score percentages which were equivalent to the independent, instructional, and frustration levels of an independent reading inventory. Bormuth concentrated on the area of comprehension rather than word recognition. He had 100 subjects write a cloze test and a multiple-choice comprehension test both of which were constructed on the same English selection. From his findings he reported that a cloze test score of 38% was equivalent to a comprehension (multiple choice) score of 75% and that a cloze score of 50% was equivalent to a comprehension score of 90%. In other words, a cloze test score of 38% represented the lower limits of a student's instructional level and the upper limits of his/her frustration level while a cloze score of 50% marked the

lower limits of his/her independent level and the upper limits of his/her instructional level. Table IV shows this information.

Similar studies have since been carried out by Bormuth (1968b, 1969) and Rankin and Culhane (1969).

In his second study, Bormuth (1968b) again compared multiple choice comprehension scores to cloze test scores based on the same material. In addition, he compared cloze test scores with Gates Oral Reading Test scores in order to get a cloze test equivalent score for word recognition. The comprehension scores and cloze test scores were similar to those reported in his earlier study. A 44% cloze score corresponded to a 75% comprehension score (the cloze score was 38% in the previous study) and a 57% cloze score corresponded to a 90% comprehension score (it was 50% in the earlier study). The oral reading scores and cloze test scores were also matched. Bormuth found that the upper limit word recognition score, that is, 98% (the independent level) corresponded to a cloze test score of 54% and that the lower level, 95% (the instructional level), corresponded to a cloze score of 34%. These findings are summarized in Table V.

On the basis of the discrepancy between the cloze test scores of 44% and of 34% each of which should represent the lower limits of a student's instructional level, Bormuth suggested that the "rule of thumb" informal reading inventory

TABLE IV

Cloze Scores and Their Informal Reading Inventory Equivalents (Bormuth, 1967a)

Equivalent Cloze Scores	Informal Reading Inventory		
	Independent	Instructional	Frustration
	above 50%	38% to 50%	below 38%

TABLE V

Cloze Test Scores and Their Individual Reading Inventory Score Equivalents (Bormuth, 1968b)

Cloze Test	Independent Reading Inventory			
	Comprehension		Word Recognition	
	Instructional 75%	Independent 90%	Instructional 95%	Independent 98%
Cloze Scores	44%	57%	34%	54%

measures might be questioned. In his 1969 study, he did just that. However, he expressed no concern for the fact that the percentages of his 1967 study were somewhat different from those of his 1968 study. Table VI points out these differences.

In his third study, Bormuth (1969) used a design involving matched pairs. One member of the pair was given a cloze test and the other member was given a multiple choice test (comprehension), treatment, and a second multiple choice test. The purpose of the experiment was to determine the amount of gain caused by the treatment and the implications this had for setting cloze test independent and instructional percentage limits. That is, if the member of the pair who took the comprehension test showed a significant gain following treatment, his partner's cloze test score was considered to be in the instructional range. Bormuth found that students scoring less than 37% on the cloze tests learned little or nothing from the treatment (i.e., they were at their frustration levels) and that students scoring more than 37% on the cloze tests tended to gain information at a constant level as a result of the treatment. From this he concluded that the 75% comprehension score of the informal reading inventory might well be lowered to 65% since this score was equivalent to the 37% cloze test score and that even those students working at their so-called independent levels (90% or more) could learn from further instruction.

TABLE VI

A Comparison of Informal Reading Inventory Comprehension Levels with Cloze Test Scores from Two Bormuth Studies

Informal Reading Inventory Comprehension Level	Cloze Test Scores	
	Bormuth 1967a	Bormuth 1968b
Instructional	38%	44%
Independent	50%	57%

See Figure 2.

The studies of Alexander (Mork, 1971) and Rankin and Culhane (1969) showed findings similar to those of the Bormuth studies. Table VII, similar to one found in Mork (1971), summarizes these results. All numbers used in Table VII are cloze score percentages.

As a result of the information in Table VII, Mork (1971) suggested that cloze test scores of about 45% could be used to indicate a student's instructional level and scores of about 60% to indicate his/her independent level. The author would be inclined to lower the instructional level figure to about 40% using the same information.

Using the above, it is clear that a teacher could (a) take a passage from any selection, (b) construct a cloze test on it, (c) administer the test to his/her pupil(s), (d) mark the tests, (e) convert the raw scores to percentages, and (f) determine the suitability of the material for the pupil(s) tested.

While the testing of this concept was not a part of the present study, it was obviously the major reason for carrying out the study.

Placement Methods Reviewed

The methods used to select suitable materials for students have been readability formulae used in conjunction with standardized test scores or informal reading inventory

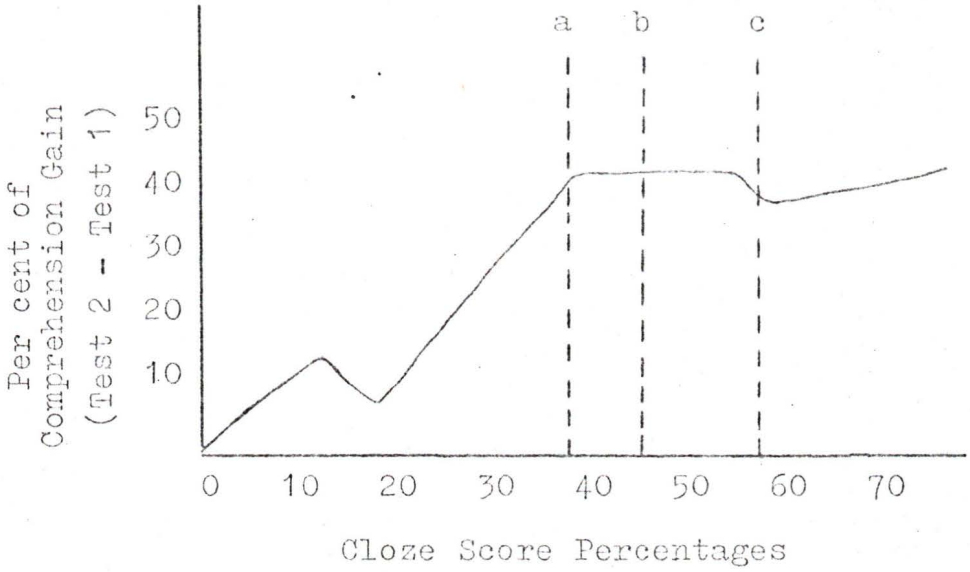


Figure 2 Comprehension scores and cloze test equivalent percentages (Bormuth, 1969).

[a]Represents a score of 65% on the first comprehension test.

[b]Represents a score of 75% on the first comprehension test.

[c]Represents a score of 90% on the first comprehension test.

TABLE VII

Cloze Test Scores Related to Independent, Instructional and Frustration Reading Levels

Study	Reading Development Level		
	Independent	Instructional	Frustration
Bormuth, 1967a	50	38 to 50	38
Bormuth, 1968b, comp.	57	44 to 57	44
Bormuth, 1968b, w.r.	54	34 to 54	34
Bormuth, 1969	57	37 to 57	37
Anderson, 1969	61	47 to 61	47
Rankin & Culhane, 1969	61	41 to 61	41

scores and cloze procedure. The author discounted the notion of referring a student to the librarian "to find a book" as being less than adequate.

Standardized Tests

The criticisms of standardized tests as placement devices have been based on the costs of such tests in terms of monetary resources; on the time required to administer, mark and interpret such tests; and on the usefulness of the information given by the tests.

Bormuth (1968b) commented on the high costs of multiple choice and sentence completion tests in terms of both monetary and human resources. Such question types, especially multiple choice items, have formed the basis of most standardized tests.

Of greater significance, however, are the questions he raised about the worth of the information given by standardized test scores. Was the score on such a test a measure of the passage difficulty as it was purported to be, or was it instead a measure of test difficulty (Bormuth, 1968b)? Bormuth (1967b) asked a further question. To what degree were multiple choice scores inflated by subjects who made "good guesses"? Skyer (1969) supported the contention that multiple choice scores could easily be inflated. Jones and Pikulski (1974) claimed that standardized tests were useful for ranking or screening students but since they tended to overestimate a student's ability they often led to placing a

student in materials which were not suitable.

Informal Reading Inventories

The utility of the informal reading inventory has also been questioned. Bormuth (1968b), Zintz (1970) and Jones and Pikulski (1974) commented positively on its validity as a placement device but each also noted the great amount of time required to administer such inventories to large classes. A further difficulty pointed out by Zintz (1970) was the lack of teachers who had been adequately trained to use informal reading inventory techniques. One other question was raised by Bormuth (1969) and mentioned earlier. Is there sufficient research to justify the percentage scores attached to the reading development levels?

One very real strength of the informal reading inventory which is not found in standardized tests or cloze procedure, is its measurement of oral reading ability and word attack skills rather than simply comprehension (Burron & Claybough, 1972).

Cloze Procedure

Cloze procedure has advantages over the other methods in that it ". . . is clearly one of the most, if not the most, convenient and widely applicable techniques ever suggested [Groff, 1971, p. 677]" for determining readability. Rankin (1959a) supported this by saying that cloze testing was a cheap method of providing ongoing evaluation. A further advantage of cloze procedure over other methods was

that ". . . the closer the test procedures and materials approximate actual teaching, the more realistic and functional will be the reading level indicated [Jones & Pikulski, 1974, p. 433]." Standardized tests have not qualified as approximating the materials of actual teaching; neither have informal reading inventories tended to use materials which were later used in instruction. "Cloze procedure does not inject an extraneous reading task into the process of determining readability [Bormuth, 1967b, p. 489]."

The use of the completed cloze test form with the subject as a means of helping both the subject and the teacher understand the errors made has helped to offset some of the advantages of the oral aspect of the informal reading inventory (Rankin, 1959a).

Readability Formulae

Many proponents of cloze procedure have compared it to readability formulae to the general disadvantage of the formulae.

The classic study showing the advantages of cloze procedure over readability formulae was done by Taylor in 1953. In this often quoted experiment, Taylor set out to show that in some cases readability formulae would not discriminate between easy and difficult prose passages. Because readability formulae relied heavily on sentence and word lengths to rate readability, Taylor hypothesized that a selection containing common words in long sentences would

be judged harder than one containing short, but uncommon, words in short sentences. He further predicted that cloze procedure would accurately rate the reading difficulty of these passages. His experiment proved him right on both counts. A passage from Gertrude Stein's Geography and Plays was rated as easy by both the Flesch and Dale-Chall Formulae while a passage from Erskine Caldwell's Georgia Boy was rated more difficult. Cloze procedure, on the other hand, ranked the passages in an order comparable to the one in which the experimenter had subjectively placed them (Taylor, 1953).

Taylor explained the success of cloze procedure by saying that it did not count the elements of English prose, as did formulae, but it did measure the effect these elements had on readability (Taylor, 1953).

Dale and Chall (1948) agreed that readability formulae were not sensitive to subtle variations of meaning. Fries (1963) added emphasis to this weakness when he said,

> Readability scales and measures of grade placement for reading materials give primary attention to vocabulary items but a language cannot be identified with a dictionary [pp. 63-64].

Groff (1971, p. 677) pointed out that a publisher, knowing how readability formulae work, could cheat since by altering a few words the readability but not the comprehensibility of a selection could be changed.

Possibly the best statement pointing out the advantages of cloze procedure over readability formulae comes from

Rankin (1969)

> It [cloze procedure] measures the suitability of the material for a particular type of reader. It takes his background into account as well as his language skills and other factors [p. 577].

Klare (1966) who has done a great deal or work with readability formulae noted that cloze procedure has helped researchers develop better readability formulae. Bormuth (1967b) claimed that cloze procedure, along with linguistics and mathematics, has helped to improve readability formulae "by up to 75% [p. 489]."

Earlier in the chapter it was pointed out that reliable readability formulae tended to be so complex and time consuming to apply that they were relatively beyond the use of the classroom teacher. The purpose of this study, of course, was to prove that cloze procedure was not.

Cloze Format

Cloze test construction has consisted of choosing a suitable passage and then, beginning with the first word of a paragraph, typing the passage. As it has been typed, various words have been deleted and replaced with blank spaces of standard lengths (usually about 15 spaces long). The end of the passage typed has coincided with the end of a sentence.

In constructing a cloze test, three factors have had to be considered: deletion rate, length of test passage, and number of test passages.

Deletion Rate

Taylor (1953) found that scores on cloze tests did not vary significantly when random deletion rates were compared with "nth" word deletion rates. He concluded that since the construction of a cloze test was easier when every "nth" word was deleted, he would use that method. Oller and Conrad (1971) supported the "nth" word deletion rate on the basis of their studies. Non-random deletion rates, that is, rates which intentionally delete word classes, have also been used. The cloze forms thus generated, however, have generally been used for some instructional purpose such as teaching students for whom English was a Second Language (Louthan, 1965).

Commonly used deletion rates have varied from one-in-five (Taylor, 1953; Bormuth 1967a) to one-in-eight (Anderson, 1966) to one-in-ten (Bloomer, 1962). There seemed to be no evidence that favoured any one of the above deletion rates over any other although Taylor (1953) reported that cloze test scores tended to converge as more successive words were taken into account. Bormuth (1965) reported similar findings. The implication of this, to the author, was that a one-in-five rate may have proven more reliable than a lower deletion rate (one-in-eight or one-in-ten) over the same material simply because the one-in-five rate would take more potential cloze units into account. Hafner (1966) found that deletion rates greater than one-in-five tended to

remove too much information.

The first purpose of this study was to find evidence to suggest an optimum deletion rate.

Cloze Test Length

The lengths of selections chosen for cloze tests have varied from 30 words (Anderson, 1966) to 300 words or more Taylor, 1953). The comments by Taylor to the effect that cloze scores tended to converge as more words were taken into account suggested that the reliability of very short cloze passages was suspect. In fact, Taylor (1953, pp. 424-425) directly stated that 35 blanks discriminated better than 16 blanks. The study of optimum sample size by Bormuth (1965, p. 115) suggested that with as few as 30 subjects, a cloze test should have more than 50 blanks in its length. As some of the figures used by Bormuth were extrapolated, exact lengths were unclear.

The second purpose of this study was to suggest an optimum cloze test length. As shorter lengths implied less teacher-time necessary for constructing, administering, and marking, the shortest length which still indicated high reliability was sought rather than simply the length which gave the greatest reliability.

Number of Selections

The number of selections tested from a single passage have ranged from one (Bormuth, 1967a) to three (Anderson, 1966). A study by Bormuth (1964) indicated that the use of

one form only might give unreliable cloze scores. He noted, also, that differences between forms decreased as the forms got longer.

The third purpose of this study was to determine how few cloze test forms could be used to test a selection and still retain a high reliability.

Administration and Marking of Cloze Tests

Cloze test administration generally has consisted of giving all subjects a copy of the test, asking them to complete as many items as possible, and allowing them sufficient time to finish (Burron & Claybough, 1972).

Cloze responses generally have been marked correct only if they were identical (with minor spelling variations) to the words deleted (Taylor, 1953). A study by Coleman and Miller (1967) gave very strong support to this practice. They found that when synonyms and word class responses were also scored as correct, the resulting raw scores correlated .99 with the exact word cloze scores. They concluded that there was little point in taking the extra marking time to include synonyms and word class forms in cloze scores. Oller (1972) did not agree with these findings. He found that exact word plus acceptable synonym scores were significantly different from exact word scores. This difference might be explained by the fact that the subjects in the Oller study were English as a Second Language students and that they would, therefore, have a far better understanding

of some word classes than they would of others. Oller's study was not replicated using students for whom English was a native language.

In the present study, the author used the administration procedures outlined earlier and the exact response method for scoring.

Interpretation of individual's scores was not a part of this study.

Reliability and Validity of Cloze Procedure

Cloze procedure has generally been found to possess high reliability. Taylor (1956) found that cloze test forms correlated .88 with other cloze test forms. Hafner (1964) reported a split-half reliability of .79 on a cloze test of 50 blanks. Bormuth (1969) found split-half reliabilities of .92 and .89.

Bormuth (1968a) claimed that cloze procedure had content validity in that, since a cloze test was constructed on the material a student would be required to read, it obviously provided a demonstration of his/her achievement in that material. However, the most often used type of validation of cloze procedure has been criterion oriented. Concurrent validity has been shown of cloze tests with comprehension tests at .70 to .88 (Taylor, 1957); of cloze tests with reading and dictation tests at .80 and .82 (Oller & Conrad, 1971); of cloze tests with standardized reading

tests at .78 and .73 (Rankin, 1959a); and of cloze scores with intelligence test scores at .73 (Hafner, 1964). The above reported findings are by no means all that are available but in the author's opinion they are representative of the data available.

It must be noted before leaving this section that a study by Kirby (1970) found a significant difference between test scores on four instruments: Gates, Gilmore Oral, Gray Oral and cloze. These differences were found to persist until subjects reached the grade five level. It must be pointed out, however, that the cloze tests Kirby used consisted of selections which ranged in word length from 23 to 212. It seemed doubtful to the author that a cloze test score based on a passage of 23 words should be considered reliable. It was interesting to note that as the cloze test passages grew longer, the significant differences disappeared. It was also interesting to note that the findings of Anderson (1966) showed that cloze tests could be used to determine readability equally well at the primary and senior secondary levels. It would be useful to have the Kirby study replicated using longer cloze test passages.

Other Uses of Cloze Procedure

A number of writers suggested that cloze tests were superior or equivalent to comprehension tests and that they could be used in place of such tests. Skyer (1969) reported

that "multiple choice did not correlate more highly with the independent measure than did cloze [p. 42]." Preston (1964), on finding that college students did better than chance in answering multiple choice comprehension questions before covering the material, resolved to use cloze tests instead to measure material learned. A study by Rothkopf (1968) used cloze in a test/re-test situation in place of other measuring devices. Dickens and Williams (1964) found an oral cloze test better than a multiple choice test for studying the comprehension of spoken messages.

Feldman (1965) fairly well summarized the use of cloze tests when he said, "The procedure has been found to be effective as a measure of effectiveness of communication, readability, intelligence, knowledge and comprehension [p. 135]."

Perhaps the most unusual use was reported in Hafner (1966) where cloze procedure was used in personality testing.

Others have used cloze procedure as a teaching device. Rankin (1969) claimed that training through cloze procedure improved reading comprehension, listening ability, writing ability, and identification of linguistic structures. Culhane (1970) reported that the use of cloze procedure led to vocabulary development, increased comprehension and a better understanding of language patterns.

Cloze procedure has obviously caught the imagination of a large number of educators.

Summary

Chapter II presented an analysis of readability measures with a special emphasis on cloze procedure. In an effort to justify the use of cloze procedure as a readability measure, its development, dynamics, utility, and strengths were discussed.

CHAPTER III

DESIGN OF THE EXPERIMENT

The material in this chapter has been divided into two parts. Part I contains a description of the instruments used and Part II contains a description of the two phases of the experiment.

PART I

Description of the Instruments

In this study three types of measures were used: cloze test scores, standardized reading test scores, and a ranking of the subjects by the teacher concerned in descending order of reading ability.

Cloze Tests

The cloze tests constructed for this study were based on selections from the fourth year basal reader, Believe and Make Believe by Sheldon and Provost (1961). This book was available as a supplementary reader in the Hope School District. The author could find no evidence to suggest that it had been used previously by any of the subjects. The book's only indication of readability was the numeral 4 on

its spine.

Selections for cloze testing were chosen because the author of this study believed they would be of high interest to fourth year pupils.

Each cloze test began with the first word of a paragraph and was continued to the end of the sentence which contained the last word of the predetermined length. Contractions and hyphenated words were counted as single words. Deletions began with the fifth, eighth, or eleventh word, depending on the particular type of deletion used.

Sample tests. The selections chosen for the sample tests were, "In Which Pooh Goes Visiting and Gets Into a Tight Place," by Milne (pp. 83-91) and, "The Story of the First Bow and Arrow," by Osborne (pp. 180-188). The sample test constructed on the Milne selection began with a paragraph approximately one-third of the way through the selection and was 181 words long. Every fifth word was deleted so that a total of 36 blanks, or potential cloze units, were created. The sample test created on the Osborne selection began with the first word of the story and was 161 words long. Again, every fifth word was replaced by a blank so that 32 potential cloze units were created.

Formal tests. For the formal cloze test forms, two further selections were used: "The Voice of Winter," by Gall and Crew (pp. 63-72) and, "Little Chuck Becomes a Friend," by Bridgemen (pp. 40-50).

Two passages were used from the first selection. The first began with the first word of the fourth paragraph and was 301 words long. Cloze test Forms A, B, and C were constructed on this passage. Form A used a one-in-five deletion rate. Altogether it contained 60 blanks. Form B used a one-in-eight deletion rate. It contained 37 blanks. Form C used a one-in-eleven deletion rate which resulted in 27 blanks.

The second passage came from near the end of the selection. It was 315 words long. Forms D, E, and F constructed on this passage contained 60, 37, and 27 potential cloze units respectively. Table VIII indicates cloze test and passage relationships.

The author had the cloze tests stapled together in sets. Set One contained Forms A and D, Set Two contained B and E, and Set Three contained C and F.

The forms described above were used in the first phase of the experiment.

In the second phase of the experiment, three further cloze test forms were constructed. They were based on three passages from the selection by Bridgeman. This selection was divided into three short chapters. Each passage used began with the first word of a chapter. Cloze test Form G was constructed on the first passage which contained 265 words. It used a one-in-five deletion rate and contained 50 blanks. Form H was based on the second passage which was 271

TABLE VIII

"The Voice of Winter"

Deletion Rate	Passage #1 301 words	Passage #2 315 words
one-in-five 60 deletions	Form A	Form D
one-in-eight 37 deletions	Form B	Form E
one-in-eleven 27 deletions	Form C	Form F

words long. It, too, used a one-in-five deletion rate and contained 50 blanks. Form I was constructed on the third passage, contained 260 words, used a one-in-five deletion rate, and had 50 blanks. Table IX shows the cloze test-passage relationship for this selection.

Standardized Reading Tests

The Gates-MacGinitie Reading Tests, Survey D, Form 2M, Separate Answer Sheet Edition (Gates & MacGinitie, 1965) were chosen to provide a test of concurrent validity.

This particular standardized test was chosen because (a) the author had used it previously, (b) it had high alternate form (.85) and split-half (.88) reliability, (c) it appeared to be one of the reputable current standardized tests then on the market (Buros, 1972), (d) it grew out of the Gates Reading Tests and was thus expected to have concurrent validity with that test, and (e) it was relatively simple to administer and score.

Survey D was selected because it had been devised for use in grades 4 to 6. Form 2M was randomly selected. It contained three sections: Speed and Accuracy (36 items), Vocabulary (50 items), and Comprehension (52 items). The Separate Answer Sheet Edition was chosen because of cost factor (it could be re-used) and because it was easy to score.

The Gates-MacGinitie Reading Tests materials which the author used were the "Technical Manual"; the "Teacher's

TABLE IX

"Little Chuck Becomes A Friend"

Cloze Test	Passage #1 265 words	Passage #2 271 words	Passage #3 260 words
one-in-five deletion rate	Form G 50 blanks	Form H 50 blanks	Form I 50 blanks

Manual"; the Gates-MacGinitie Reading Tests, Survey D, Form 2M; "IBM 805 Answer Sheets"; and the "IBM 805 Scoring Key" (Gates & MacGinitie, 1965).

Teacher's Judgement

The author used the teacher's judgement of her pupils' reading abilities to provide a further test of concurrent validity of the cloze procedure.

In the same time period as the cloze tests Forms G, H, and I and the Gates-MacGinitie tests were given, the classroom teacher was asked to rank her pupils according to her perception of their reading abilities. She was told that she could use any objective or subjective measures she wished but that her final ranking should not depend on any one measure. Further, it was made clear that the rank-ordering must not depend in any way on her pupils' results on either the cloze or standardized tests. To offset this possibility as much as possible, the teacher was asked to administer the tests and return them immediately afterward to the author.

It was explained to the teacher that if she were unable to rank two or more of her pupils in relation to one another, she should give them both (or all) the same rank.

In Part II, the administration and use of these measures is discussed.

PART II

The experiment was run in two phases. The first phase took place during the second week of April, 1973, and the second phase during the second and third weeks of November, 1973. Since the two phases were distinctly different this section has been divided to treat them separately.

Phase I

The Sample

Fourth year pupils in two schools in the Hope School District were involved in Phase I. In one school, the total fourth year population was 82; in the other it was 35. In the first school, pupils had been grouped homogeneously. They were in three classes. Two classes contained only fourth year pupils. The other contained both fourth and fifth year pupils. In the second school, the subjects formed a heterogeneous group. In all, four teachers were involved with the administration of the tests: three in the first school and one in the second.

Selection of the Data

Three groups were needed for the first phase of the study. The pupils in the first school were divided into two groups and the total population of fourth year pupils in the second school formed the third group. To insure a degree of heterogeneity in the first two groups, each teacher in the first school was given enough tests from Set One for half of

his/her pupils and enough of Set Two for the other half. The pupils in the second school were given Set Three.

The number of paired tests in each set were, therefore, as follows: Set One, 41; Set Two, 41; Set Three, 35. One of the pairs in Set Three had only one completed form which reduced the total number in the set to 34. Seven paired tests from each of Sets One and Two were then randomly withdrawn so that all sets contained 34 paired forms. Table I shows the relationship among tests and subjects.

Testing Procedures

Sample tests. Each teacher was asked to administer the two sample tests to his/her class and to mark and discuss the tests with the class. Marking keys were provided with the sample tests. One purpose for this was to avoid a practice effect when the formal testing took place. The other purpose was to insure that all participants--students and teachers--fully understood the cloze task before the formal tests were administered.

Formal tests. Teachers were then asked to set aside the period immediately following recess on two successive days for the purpose of testing. They were asked to choose days in the middle of the week to eliminate as far as possible the chances of high absenteeism. Directions given to the teachers regarding the actual testing procedure were as follows: (a) distribute the sets of cloze tests to all pupils (do the first test one day and the second the next);

(b) insist on quiet; (c) allow all pupils the opportunity to complete the test; (d) collect the tests.

Once their classes had completed the two formal test forms, teachers returned them to the author. Marking keys were not provided for the formal tests.

Marking. All forms were marked and scored by the author using the exact answer scoring method. Results were re-checked to insure accuracy.

Analysis of the Data

Cloze test scores were tabulated by the author and processed at the Computing Centre of the University of Victoria. Calculations with the resulting coefficients were done by the author.

Deletion rate. To test the reliability of the three deletion rates, a Pearson product-moment correlation was found for the paired results within each of the three sets. The resulting correlation coefficients were tested for individual levels of significance (Popham, 1967). Following that, each was compared with the other two and tested for significant differences using the procedure outlined by Ferguson (1966) for finding the "significance of the difference between two correlation co-efficients for independent samples [pp. 153-154]."

Following the statistical treatment of the data, no one particular deletion rate showed as being the best. As a consequence, practical considerations relating to the

deletion rate were weighed.

As a result of these deliberations, the one-in-five deletion rate was accepted by the author as the optimum deletion rate for the cloze format. These reasons are discussed in Chapter IV. A decision regarding the optimum deletion rate had to be made at this point because subsequent parts of the study were dependent on this first part.

Test length. The test scores for Set One(Forms A and D) were re-tabulated in seven categories. These categories were as follows: (a) Category One matched the scores over the first 30 words (6 blanks) of Form A with the scores of the first 30 words of Form D, (b) Category Two matched the scores over the first 50 words (10 blanks) of the two forms, (c) Category Three matched the scores for the first 100 words (20 blanks), (d) Category Four matched the scores of the first 150 words (30 blanks), (e) Category Five matched the scores of the first 200 words (40 blanks), (f) Category Six matched the scores over the first 250 words (50 blanks), and (g) Category Seven matched the scores over the 300 words (60 blanks). Table X indicates the categories created from cloze test Forms A and D.

A Pearson product-moment correlation was found for the matched scores in each of the seven categories. The resulting correlation coefficients were tested for individual significance (Popham, 1967). Following this, each coefficient was statistically treated using the Spearman-Brown

TABLE X

Category Tabulations of Cloze Test

Form A and Form D Scores

Cloze Test	Category						
	1	2	3	4	5	6	7
Form A	30 words	50 words	100 words	150 words	200 words	250 words	300 words
Form D	30 words	50 words	100 words	150 words	200 words	250 words	300 words

formula (Ferguson, 1966) to make all coefficients representative of tests of the same length. This procedure was necessary before the coefficients could be compared for statistical differences. Using the procedure outlined in Ferguson (1966) for testing the differences of correlation coefficients from correlated samples, the resulting coefficients were tested.

Since the statistical analysis of the data again did not point out the best cloze test length, a consideration of the practical aspects of cloze test length was made. On the basis of this, a discussion of which takes place in Chapter IV, the author decided to use a cloze test length of 250 words. This decision was necessary before proceeding to the testing which took place in the second phase of this study.

Phase II

This phase, as mentioned earlier, took place seven months after Phase I. This time was required to process the data of Phase I on which results the construction of cloze test Forms G, H, and I depended. It was also desirable, in terms of the teacher's judgement measure, to give the classroom teacher an adequate amount of time to get to know her pupils.

The Sample

The sample for Phase II of the study consisted of one class containing 33 fourth year pupils from a third school

in the Hope School District. Since the class consisted of all the fourth year pupils in the school, it was heterogeneous. As the number of tests, and therefore the amount of time required to administer them, was greater, this phase of the study experienced a higher experimental mortality than did the first. Three pupils did not complete all of the tests so in fact the number of subjects taking part in the study was reduced to 30.

Selection of the Data

From each of the 30 students participating in this phase of the experiment five scores were obtained. They were the cloze test scores on Forms G, H, and I; the standardized reading test score on the Gates-MacGinitie Reading Tests (1965); and the rank-order score based on the teacher's judgement of reading ability.

Testing Procedures

Sample tests. Sample tests were provided to the teacher with the same instructions as those given in Phase I. As the author was in the same school, he spent part of one period discussing Sample Test 2 with the class.

Formal tests: cloze. Following the administration and discussion of the sample tests, cloze test Forms G, H, and I were administered on successive days. The timing and directions for the administration of these were the same as those in Phase I. On completion of each test form, it was returned, unmarked, to the author.

Formal tests: Gates-MacGinitie. Once the three cloze test forms were completed, the classroom teacher administered the Gates-MacGinitie tests. The administration of the tests took two days with the "Speed and Accuracy" and "Vocabulary" sections being given one day and the "Comprehension" section the next. Again, the testing time was the period immediately following recess. The directions set out in the "Teacher's Manual" for the Gates-MacGinitie Reading Tests (Gates & MacGinitie, 1965) were followed. Once these tests were completed, they were returned, unmarked, to the author. The teacher's judgement (rank-ordering) was given to the author at the same time as were the Gates-MacGinitie tests. At no time during the formal tests did the author visit the class.

Marking. These tests were marked in the same way as were those in Phase I.

Analysis of the Data

Data was again tabulated by the author and processed at the Computing Centre of the University of Victoria. Calculations with the resulting coefficients were done by the author.

Number of tests. In the first series of tests, a Pearson product-moment correlation was used to compare the score of the correct answers to the odd questions on cloze test Form G with the score of the correct answers to the even questions on the same form. Similarly, a Pearson product-moment correlation was found for the cumulative

split-half scores on Forms G plus H, then on Forms G plus H plus I. These coefficients were treated by the Spearman-Brown formula to find a whole-test reliability estimate (Ferguson, 1966). They were then tested for individual significance (Popham, 1967).

Following this, the corrected coefficients were inspected for significant differences. It had been intended to test the differences between the coefficients statistically, but as is reported later, this was not possible.

As no statistical evidence could be found to favour one, two, or three tests per selection, a decision on this factor was determined on other bases which will be related in Chapter IV. The decision made, which was necessary before proceeding to the next series of tests, was that only one test per selection was necessary.

Reliability. The second series of tests made on the data found the Pearson product-moment correlation coefficients of cloze test Forms G and H, G and I, and H and I. These coefficients were compared using Ferguson's test to compare the significance of the difference between two correlation coefficients for the same sample (1966).

The sets of data found above were used to determine the internal consistency of the cloze test format.

Validity. The third series of tests were designed to measure the concurrent validity of the cloze test format determined with, first, the Gates-MacGinitie Reading Tests

(1965) and, second, the teacher's judgement. Accordingly, a Pearson product-moment correlation was computed which matched the cloze test scores from Forms G, H, and I with the total scores from the Gates-MacGinitie Reading Tests (1965). This coefficient was tested for significance (Popham, 1967). Next, a Spearman Rho was found which compared the rank-order of the cloze test scores from Form G with the rank-order of the teacher's judgement. The resulting coefficient was tested for significance (Popham, 1967).

Finally, to provide a frame of reference for the Spearman Rho calculated above, a second Spearman Rho was calculated, this time comparing the rank-order of the Gates-MacGinitie scores with the rank-order of the teacher's judgement. This coefficient was tested for significance (Popham, 1967).

Summary

This chapter included a description of the instruments used in the study, an outline of the sample, an account of the testing procedures, and a discussion of the statistical procedures used on the data.

CHAPTER IV

ANALYSIS AND RESULTS

This chapter presents an analysis of the data obtained from the three types of measures outlined in Chapter III: cloze tests, standardized tests, and teacher's judgement. It discusses the statistical analyses carried out. In addition, it relates the practical considerations relating to test simplicity which were taken into account in the development of the cloze test format. Finally, it reports the procedures used to test the hypotheses.

Procedures for Developing Cloze Test Format

Before testing the hypotheses, it was necessary to develop a cloze test format. This development occurred through answering the three questions posed in Chapter I. The nature of the present study was such that each question had to be dealt with fully before proceeding to the next question. This meant that all the data relating to a particular question had to be considered at the same time. In dealing with this data the following procedure was used. First, the statistical data were analysed. Any decisions that were possible, based on the significance or significant

differences of this data, were made. Second, where the statistical analysis failed to discriminate between significant findings, a further analysis was tried without models. For example, if there was no statistically significant difference between correlations of .7 and .9, the difference was analysed in terms of the trend it seemed to show. Third, when as a result of the statistical analyses no decision could be made, practical considerations weighted in favour of ultimate test simplicity were the deciding factors.

At all times during the development of the format, the author was very much aware that choosing format items with low reliabilities would likely cause the rejection of both hypotheses.

Analysis of the Development of Cloze Test Format

Question One

Of the deletion rates one-in-five, one-in-eight, and one-in-elevent, which was the maximum rate which still possessed satisfactory reliability?

Statistical considerations. In order to determine the maximum deletion rate, three groups of pupils were each given one set of cloze test forms. Pearson product-moment correlations were found for the two cloze test forms in each of the three sets. Table XI shows these correlation coefficients.

TABLE XI

Correlations of Cloze Test Forms

Within Sets One, Two, and Three

Cloze Set	Deletion Rate	Group	Correlation coefficient	Coefficient needed to be significant[a]
One (Forms A,D)	1-in-5	1[a]	.838*	.436
Two (Forms B,E)	1-in-8	2[a]	.791*	.436
Three(Forms C,F)	1-in-11	3[a]	.668*	.436

* $p < .01$

[a] $N = 34$

Having determined that all correlation coefficients were significant at the .01 level (Popham, 1967), the significance of the difference between each of the coefficients was tested, using the model outlined by Ferguson (1966) to test the significance of the difference between two correlation coefficients for independent samples. The formula for this test is as follows

$$Z = \frac{Zr_1 - Zr_2}{\sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-3}}}$$

where N stands for the number of subjects and Zr for the normalized correlation coefficient obtained by using the Fisher Zr transformation (Ferguson, 1966).

As was pointed out in Chapter I, this test was not intended to be used in quite this way. While the samples were independent, the tests for which the correlation coefficients were calculated were not identical. That is, although in each case exactly the same passage was cloze tested, the differing deletion rates caused different words to be deleted on each test form. Bearing this limitation in mind, the formula was applied. Table XII gives the results of the applications of the formula.

For Z to be significant at the .05 level of confidence, it would have had to have been 1.96 or larger (Ferguson, 1966). As can be seen from Table XII, none of the coefficients differs significantly from any other. To determine

TABLE XII

Significance of the Differences Between the Correlation Coefficients of Sets One, Two, and Three

Correlations compared	Differences	Z	Magnitude of Z to be significant at .05
$r_1 - r_2$	.150	.591	1.96
$r_2 - r_3$	.263	1.035	1.96
$r_1 - r_3$	.413	1.626	1.96

deletion rate, therefore, a further analysis had to be made. It should be noted, however, that the difference between the correlations representing Sets One and Three was significant at the .104 level of confidence for a two-tailed test and at the .052 level for a one-tailed test.

It seems obvious that, other factors such as sample size and test length being equal, the higher the correlation coefficient is between two forms, the more reliable are the forms. Ferguson (1966) explained why this is so. He stated that a correlation coefficient does not represent the portion of overlap of two measures. Rather, he said, the square of the correlation coefficient represents the percentage of overlap. That is, a correlation coefficient of .8 does not represent twice as much overlap as does one of .4; it represents four times as much overlap. Table XIII indicates the degree of overlap shown by the correlation coefficients from Sets One, Two, and Three.

While the author was unable to find any criteria existing in this area, he determined that any coefficient which predicted less than a 50% overlap would not be acceptable. Using this criterion, Set Three, the one-in-eleven deletion rate, was eliminated.

Practical considerations. Ferguson (1966, p. 381) went on to state that reliability is a function of test length. That is, other things being equal, a longer test is more reliable than a short one. The author surmised that while

TABLE XIII

Degree of Overlap of the First Form with the Second In Each of the Three Cloze Test Sets

Set	r	r^2	% of Overlap
One	.838	.702	70.2
Two	.791	.626	62.6
Three	.668	.446	44.6

the cloze tests in Sets One and Two covered exactly the same material, Set One was longer since it contained 60 blanks while Set Two contained 37 blanks. The author further surmised that, since Question Two was to test minimum cloze test length, the possibility of reducing test length was greater by beginning with 60 blanks and a higher reliability than by beginning with 37 blanks and a lower reliability. Based on these considerations, the deletion rate of Set One, one-in-five, was adopted. The reliability of this deletion rate over two forms of 300 words each was, as noted earlier, .838.

Question Two

Of the cloze test lengths of 30 words, 50 words, 100 words, 150 words, 200 words, 250 words, and 300 words, which was the shortest length that still possessed satisfactory reliability?

Statistical considerations. To find the reliable minimum length, correlation coefficients were found between the cloze test forms of Set One for the scores over equal numbers of words in each. These numbers ranged from 30 to 300 words. Table XIV indicates the correlations found.

As in Table XI, the correlation coefficient had to be greater than .436 to be significant at the .01 level of confidence when 34 subjects were used.

From Table XIV it can be seen that cloze test lengths of less than 150 words were eliminated as being less than

TABLE XIV

Correlations Over Varying Lengths of Cloze Tests from Set One

Cloze Tests	Number of Words						
Set One[a]	30	50	100	150	200	250	300
Forms A, D	.175	.369	.228	.525*	.727*	.802*	.838*

* $p < .01$

[a] $N = 34$

reliable.

Following this, the author attempted to determine whether any of the remaining correlation coefficients were significantly different from any other(s). It should be noted that no procedure exists to test such correlation differences directly (Ferguson, 1966). The coefficients were, therefore, re-calculated as if all tests had been 300 words long to begin with. This re-calculation was done using the Spearman-Brown formula which was designed for this purpose (Ferguson, 1966). This formula is as follows

$$r_{dl} = \frac{kr_x}{1 + (k-1)r_x}$$

where r_{dl} is the correlation of the test of the desired length, k is the number required to multiply the present length to get the desired length, and r_x is the correlation of the present length. The re-calculated correlation coefficients are shown in Table XV.

The differences between these correlation coefficients were then tested using the procedure outlined by Ferguson (1966) to test the significance of the differences between two correlation coefficients for correlated samples. The use of this procedure in this case varied somewhat from that in Ferguson. While the samples were correlated, the test was designed to test the differences when three correlation coefficients existed and where the three were measures of the correlations of three different tests. In this case,

TABLE XV

Projected Correlation Coefficients for Cloze Test Lengths of 150, 200, 250, and 300 Words

Correlations	Length in Words			
	150	200	250	300
present r	.525	.727	.802	.838
[n]projected r	.689	.799	.829	.838

[n] as if all tests were 300 words long.

only two tests were given and four correlations were tested. Since in all cases the correlations were somewhat interrelated in that they were all based on the same tests (Forms A and D) over differing length, it was thought the procedure could be used. The third correlation coefficient, necessary for the calculation of the formula, was taken to be the higher of the two correlation coefficients not being tested for significance. This was done in order to get higher t values and increase the possibility of discriminating between coefficients. The formula used was as follows

$$t = \frac{(r_1 - r_2)\sqrt{(N-3)(1+r_3)}}{\sqrt{2(1-r_1^2-r_2^2-r_3^2+2r_1r_2r_3)}}$$

where r_1 stands for the first coefficient, r_2 the second, and r_3 the third, and where N is the number of subjects. Results of these tests are reported in Table XVI.

Using Table B from Ferguson (1966) it was found that to be significant at the .05 level of confidence, t had to be 2.04 or greater. As can be seen from Table XVI, the 150 word length could be eliminated as being significantly less reliable than the longer lengths. Differences between the three longer lengths were not significant even at the .20 level of confidence.

There being no further statistical tests which could be applied to further limit the test lengths to be considered, the author tried another method.

TABLE XVI

Significance of the Differences Between Correlation Coefficients of Various Cloze Test Lengths

Length	Projected r	Differences over Lengths					
		150	t	200	t	250	t
150	.689	-	-	-	-	-	-
200	.799	.110	1.79	-	-	-	-
250	.829	.140	2.45*	.030	.55	-	-
300	.838	.150	2.62*	.039	.72	.009	.17

* $p < .05$

Returning the remaining correlation coefficients to their original forms the degree of overlap predicted by each was calculated. This overlap prediction is shown in Table XVII.

This table shows that all correlations predicted an overlap of more than 50% so none could be eliminated on that basis. Returning to the premise that greater length means greater reliability, the temptation was to use the 300 word length. However, other factors had to be considered.

Practical considerations. Although a test length of 1000 words would probably be more reliable than one of 200 words, it seemed obvious that the time and effort required to prepare, administer, and mark such a test would simply not be worth the gain. Not only would it be difficult to convince teachers to make such a commitment, it would also be unreasonable to expect subjects to do the extra work. The 300 word length took most pupils from 20 to 40 minutes to complete. Doubling or tripling the test length would certainly double or triple the amount of time required for subjects to write the test. Such an increase might well lead to fatigue and carelessness so that while test reliability might be increased, almost certainly a decrease in the accuracy and effectiveness of the test as a placement device would also follow.

While the 300 word length was tempting, and while the length of time it required of subject and teacher was not

TABLE XVII

Correlation Overlaps Predicted: Cloze Test Lengths

Length	r	r^2	% Overlap
200	.727	.529	52.9
250	.802	.643	64.3
300	.838	.702	70.2

prohibitive, it had two major disadvantages. First, it was not possible to type a 300 word passage on one sheet of paper without causing a great deal of crowding which could easily confuse the subject. The need for two sheets of paper obviously doubled the cost and increased, disproportionately, the marking time required since pages must be turned. Second, the resulting cloze test score was out of 60 (60 blanks in 300 words). To be useful, a cloze test score must normally be converted to a percentage. Conversion of a score out of 60 to a percentage required multiplication by the improper fraction 100/60 or 5/3. Since calculation errors could readily occur when using this multiplier, the author believed this to be a further disadvantage.

When considering the 250 word length, both of these disadvantages became advantages. Two hundred and fifty words, double-spaced, fitted comfortably on one page. To calculate a percentage score when the cloze score was out of 50 became simply a matter of multiplying by two.

Given these advantages, plus the fact that the correlation coefficient was still satisfactorily high over the 250 word length, the author determined to use it in preference to the 300 word length.

Having made a concession to practicality, it seemed in order to make one to length. The 200 word length was therefore discarded. It should be noted, in addition, that conversion of a cloze test score out of 40 (200 word length)

contained some of the disadvantage of conversion of a score out of 60. Further, it seemed simpler to make one cloze test 250 words long than to make two each of which was 200 words long. The author was concerned that a shorter length might so lower the split-half and alternate form reliabilities of the total cloze format developed that two tests would be necessary.

Thus far, the cloze format determined had a deletion rate of one-in-five and a length of 250 words.

Question Three

Of the numbers of cloze tests per passage, 1, 2, and 3, which was the lowest number which still possessed satisfactory reliability?

Statistical considerations. To determine the minimum number of cloze tests to be used, three forms, G, H, and I, were constructed on three passages from one selection. These forms were administered to the fourth group, marked, and tabulated. Two scores were found for Form G. These were the total of the correct answers for the odd questions and the total of the correct answers for the even questions. Similar scores were found for Forms H and I. These scores were then compared on a cumulative basis. That is, the odd scores from G were compared with the even scores from G, the odd scores from G plus H were compared with the even scores from G plus H, and the odd scores from G plus H plus I were compared with the even scores from G plus H plus I.

Correlation coefficients were then calculated on these comparisons. These coefficients were corrected for test lengths using a special case of the Spearman-Brown formula (Ferguson, 1966)

$$r_w = \frac{2\ r_h}{1 + r_h}$$

where r_h is the reliability for half the test and r_w is the reliability for the whole test. It must be remembered that since the correlation coefficient first calculated really only accounted for half the length, this coefficient had to be corrected. Table XVIII reports both the uncorrected and corrected correlation coefficients along with their significances.

All correlations were found to be significant at the .01 level of confidence (Popham, 1967).

The correlation coefficients were then projected so that they all represented the same length, namely 750 words. The results of this are shown in Table XIX.

It had been initially planned to test the projected coefficients for significant differences. However, as can be seen from Table XIX, the projected correlation for the shortest length was higher than the other two and in fact the longest length had the lowest correlation coefficient. Under these circumstances, when longer test lengths should mean greater reliability, it seemed absurd to pursue any calculation which might have shown the opposite.

TABLE XVIII

Number of Cloze Tests: Reliability Coefficients

Cloze Tests	Length	Correlations		
		r	Corrected r	Significant r[a]
G	250	.814	.897*	.463
G H	500	.873	.932*	.463
G H I	750	.911	.953*	.463

* $p < .01$

[a] N = 30

TABLE XIX

Projected Correlation Coefficients Over One, Two, and Three Cloze Tests

Cloze Test	Length	Projected Length	Corrected Correlation	Projected Correlation
G	250	750	.897	.963
G H	500	750	.932	.954
G H I	750	750	.953	.953

The statistical analysis therefore yielded the information that all three coefficients were reliable and that no one was significantly better than any of the others.

Using the corrected correlation coefficients, their degrees of overlap prediction were tested. These results are noted in Table XX.

As can be readily seen from Table XX, all correlations predicted overlaps of greater than 50%.

Practical considerations. In terms of teacher time and student time, clearly one test was better than two or three. Since there was no statistical evidence to support any one number of tests over the others, the author decided to use one test.

Testing Cloze Test Format

The cloze test format determined by the author was one test, 250 words long, with a one-in-five deletion rate. Each of the cloze tests, Forms G, H, and I, represented this cloze format.

The remainder of this study tested the reliability and validity of the Cloze Format developed.

In testing these factors the author encountered two problems. First, once a correlation coefficient was found, it was tested for significance. Since with an N of 30 or 34 any Pearson product-moment correlation coefficient of .456 or higher was significant, this test did not prove valuable

TABLE XX

Correlation Overlaps Predicted: Number of Cloze Tests

Test	Corrected Correlation	Correlation Squared	Percentage of Overlap
G	.897	.805	80.5
G H	.932	.869	86.9
G H I	.953	.908	90.8

as a discriminatory device. The author was left with the information that all of the coefficients were significant but with no means of concluding that one was better than another. This raised the second problem. As was apparent in dealing with the three questions, there are very few methods of testing the significance of the difference between two correlation coefficients. Since these methods only have validity in very specific instances, often the author was faced with the need to improvise. Even with the application of one of these methods, in most cases, no significant difference was found. The author was, therefore, in the position of having one or more significant correlation coefficients but having no statistical means to distinguish among them.

Therefore, the author tried another approach. Correlations found in similar situations for the Gates-MacGinitie Reading Tests (1965) were used as bases for comparison. The remainder of this section deals with the statistical and comparative procedures used to determine the values of the reliability and validity coefficients found for the Cloze Format.

Hypothesis I

The cloze test format developed for fourth year pupils has split-half and alternate form reliability.

Split-half reliability. Earlier it was shown that the corrected split-half reliability of Cloze Format was .897

and that this correlation was significant at the .01 level of confidence. Beyond this, there was no way to statistically treat the coefficient.

Therefore, the author compared this reliability figure to the split-half reliability figures given in the Gates-MacGinitie Reading Tests "Technical Manual" (1965). This publication reported the corrected split-half reliabilities at the grade four level of its vocabulary section (50 items) at .88 and of its comprehension section (52 items) at .94. The Gates-MacGinitie figures were established on a population of about 800 pupils (1965) which tends to make their correlation coefficients even better. On the other hand, Gates-MacGinitie is a speeded, or timed, test which tends to reduce the value of a split-half correlation (Ferguson, 1966). These two factors were considered to balance, which meant that the split-half reliability of Cloze Format compared favourably to those of the Gates-MacGinitie tests. Table XXI shows this comparison.

Alternate form reliability. To test the alternate form reliability of Cloze Format, correlation coefficients were computed for Forms G with H, Forms G with I, and Forms H with I. Table XXII shows the correlations so computed.

All correlations were found to be significant at the .01 level of confidence when the number of subjects was 30.

Using the model outlined in Ferguson (1966) the differences between these coefficients were tested for significance.

TABLE XXI

Cloze Format and Gates-MacGinitie Split-Half Reliabilities

Test	Grade Level	No. of Items	Reliability
G-M. Vocab.	4	50	.88
G-M. Comp.	4	52	.94
Cloze Format	4	50	.897

TABLE XXII

Alternate Form Reliability: Cloze Test Forms G, H, I

Forms Correlated	Correlation	Correlation to be Significant[a]
G, H	.958*	.463
G, I	.960*	.463
H, I	.965*	.463

* $p < .01$

[a] $N = 30$

The formula used was as follows

$$t = \frac{(r_{12}-r_{13})\sqrt{(N-3)(1+r_{23})}}{\sqrt{2(1-r_{12}^2-r_{13}^2-r_{23}^2+2r_{12}\,r_{23}\,r_{13})}}$$

where each "r" identifies one of the three correlation coefficients and where N is the number of subjects.

It is worth noting that this test was designed specifically for use in a situation like this. The results of the tests are summarized in Table XXIII.

As can be seen from the second column of Table XXIII, the differences were not significant even at the .05 level (nor were they at the .20 level of confidence). This led the author to conclude that there was consistency among the alternate cloze test forms.

Cloze Format alternate form reliability was also compared to alternate form reliability of the Gates-MacGinitie Reading Tests as shown in the "Technical Manual" (1965). Table XXIV shows this comparison.

Again, the larger N used in calculating the alternate form reliability of the Gates-MacGinitie Reading Tests (1965) tends to give greater significance to these scores. However, the fact that the Cloze Format alternate form reliability coefficients were in all cases higher caused them to compare favourably.

Using statistical and comparative analyses, Cloze Format was shown to have split-half and alternate form

TABLE XXIII

Significance of the Differences Between Coefficients of Alternate Close Test Forms

Correlation Differences		t	Significant t[a]
(H - G)	.007	.657	2.045*
(H - I)	.005	.469	2.045*
(I - G)	.002	.188	2.045*

* $p < .05$

[a] $N = 30$

TABLE XXIV

Comparison of Alternate Form Reliability:

Cloze Format and Gates-MacGinitie

Test	Correlation	Test Length
G-M. Vocab.	.850	50
G-M. Comp.	.830	52
G-M. Speed	.800	36
Cloze G-H	.959	50
Cloze G-I	.960	50
Cloze H-I	.965	50

reliability. Hypothesis I was accepted.

Hypothesis II

The cloze format developed has concurrent validity.

In this study, concurrent validity claimed was based on the comparison of two measures with Cloze Format: the Gates-MacGinitie scores and the teacher's judgement. The author was aware that the claim for concurrent validity depended on the assumption that both the Gates-MacGinitie Reading Tests (1965) and the teacher were able to measure reading ability. He was also aware that others might not share this assumption.

Cloze Format scores were compared with the Gates-MacGinitie Reading Tests (1965) scores and with the ranking which resulted from the teacher's judgement. To obtain a measure to compare with the Cloze Format-teacher's judgement correlation, a correlation was also found for the Gates-MacGinitie scores and the teacher's judgement.

The Pearson product-moment correlations of Cloze Format with Gates-MacGinitie are shown in Table XXV. In all cases, the number of items on the Gates-MacGinitie tests totalled 134 and on Cloze Forms, 50.

In all cases the correlations were significant at the .01 level of confidence. Accepting that the Gates-MacGinitie Reading Tests (1965) measured achievement in comprehension, vocabulary, and speed (Buros, 1972) which are considered to be aspects of reading ability, the magnitude of the correla-

TABLE XXV

Correlations of Clone Format with Gates-MacGinitie

Comparison	Correlations	Significant r^a
G-M - Form G	.945*	.463
G-M - Form H	.946*	.463
G-M - Form I	.952*	.463

* $p < .01$

[a] $N = 30$

tions of Cloze Format with these tests suggested that it, too, measured reading ability.

The second test of concurrent validity involved the comparison of the teacher's judgement and the rank-ordering of the scores from the two other measures. Table XXVI compares the Spearman Rhos computed by using this data.

All correlations were found to be significant at the .01 level of confidence (Popham, 1967). The author believed that the teacher, through an intimate knowledge of her pupils, could judge their reading abilities with considerable accuracy. Working from this belief it was clear that both the Gates-MacGinitie Reading Tests (1965) and Cloze Format provide measures of this ability. The author is aware of no means of testing the significance of the difference between the correlations relating to the teacher's judgement with the Gates-MacGinitie and the teacher's judgement with the Cloze Format. The fact that Cloze Format correlated more highly with the teacher's judgement than did the Gates-MacGinitie Reading Tests (1965) might indicate that Cloze Format measures reading ability better than does this particular standardized test.

The possibility raised by the last statement has been considered to be more than a possibility by some reading authorities. Bormuth (1966) claimed that cloze procedure had content validity because to perform on such a test a subject had to read. Since reading ability was what a cloze

TABLE XXVI

Spearman Rho Correlations: Cloze Format and Gates-MacGinitie Compared with Teacher's Judgement

Comparison of Teacher's Judgement and	Spearman Rho	Significant Rho[a]	t	Significant t[a]
Gates-MacGinitie	.821*	.432	7.620*	2.756
Cloze Format (G)	.925*	.432	12.888*	2.756

* $p < .01$

[a] $N = 30$

test was supposed to measure, the degree of reading ability possessed by the subject would dictate how well he/she did on the test. This claim of content validity was supported by Jones and Pikulski (1974) when they stated that cloze procedure approximated the reading task.

Based on the foregoing proofs of concurrent validity, Hypothesis II was accepted.

Summary

This chapter reported the development of Cloze Format and the reasoning that was used to justify this development. Next, it reported the analyses and procedures used to test the reliability and validity of the format developed. Finally, it reported the disposition of the hypotheses.

CHAPTER V

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

In presenting a summary of the study, this chapter reviews the problem, the purpose, and the design. Conclusions are drawn on the basis of the analyses which were reported in Chapter IV. Implications raised by the study are discussed.

Summary

One of the major problems faced by a reading instructor is the matching of the reading ability of his/her pupils with the readability of available print materials.

Beginning with the assumption that cloze procedure provides a suitable tool for making such a match, the author set as his purpose the development of a simple, reliable, and valid cloze test format. The particular aspects of cloze test format studied were deletion rates, cloze test lengths, and number of tests required to measure the readability of a selection.

Using selections from a fourth grade supplementary reader, nine cloze test forms were constructed by the author to test these three aspects of cloze test format. These

tests, along with one form of the Gates-MacGinitie Reading Tests, Survey D (1965) were administered to four groups of fourth year pupils in three schools in the Hope School District. The study was conducted in two phases. In the first phase, three groups in two schools each wrote two matched cloze test forms. From each group, 54 pairs of forms were selected. In one case, this was the total number of paired forms available, in the other two, the paired forms were randomly selected. The correlation coefficients which resulted from comparing these forms were tested for significance and for significant differences between pairs of coefficients. On the basis of statistical analyses and practical considerations, a maximum deletion rate was determined.

The set of matched forms which used this deletion rate then had correlations computed on it at lengths ranging from 30 to 300 words. The resulting coefficients were tested for significance and for significant differences. The results of these analyses, along with practical considerations relating to test length, were used to determine the reliable minimum cloze test length.

In the second phase of the study, the fourth group of students from the third school wrote three cloze test forms based on passages from one selection. This group also wrote "Form 2M, Survey D" of the Gates-MacGinitie Reading Tests (1965). In addition, the teacher ranked the group in terms

of their individual reading abilities. In all, 30 subjects participated in Phase II. The odd and even answers on the cloze test forms were correlated at three intervals on a cumulative basis. The resulting split-half correlation coefficients were corrected to represent whole test lengths and then tested for significance and for significant differences. In conjunction with practical considerations, these analyses were used to determine the reliable minimum number of cloze test forms to be used to test one selection.

The three aspects of cloze test format having been determined, the resulting format was tested for reliability and validity.

Reliability was tested by finding correlations for split-half and alternate forms. These coefficients were tested for significance and compared with similar reliabilities attributed to the Gates-MacGinitie Reading Tests, "Survey D" (1965).

The format was tested for concurrent validity by calculating the correlation coefficient of the cloze test scores with the Gates-MacGinitie Reading Tests (1965) scores. This coefficient was tested for significance. Following this, rank-order correlations were found for the Gates-MacGinitie Reading Tests (1965) and the teacher's judgement and for the cloze tests and the teacher's judgement. These coefficients were tested for significance and a comparison of the two was made.

A brief review of the literature was made to show content validity. The findings are reviewed in the next section.

Conclusions of the Study: Questions

Question One

Of the deletion rates one-in-five, one-in-eight and one-in-eleven, which was the maximum rate which still possessed satisfactory reliability?

The maximum deletion rate, determined by this study, is one-in-five.

Question Two

Of the cloze test lengths of 30 words, 50 words, 100 words, 150 words, 200 words, 250 words, and 300 words, which was the minimum length that still possessed satisfactory reliability?

The minimum reliable length, determined by this study, is 250 words.

Question Three

Of the numbers of cloze tests per passage, 1, 2, or 3, which was the lowest number which still possessed satisfactory reliability?

This study showed that one cloze test is sufficient to reliably test a selection.

Cloze Format

Based on statistical analyses and practical considera-

tions, the simple, reliable cloze format developed by this study is one test, 250 words long, with a one-in-five deletion rate.

Conclusions of the Study: Hypotheses

Hypothesis I

The cloze test format developed for fourth year pupils has split-half and alternate form reliability.

Hypothesis II

The cloze test format developed has concurrent validity as a test of reading ability.

Review of the Findings

The cloze test format developed was shown to have split-half and alternate form reliability based on statistical analyses and on comparisons with reliabilities established on a recognized standardized test.

The cloze format developed was shown to have concurrent validity based on statistical comparisons of it with a recognized standardized reading test and with the teacher's judgement of reading ability. Opinions of reading authorities were used to show that cloze procedure has also been claimed to have content validity.

Limitations

1. In this study, the small sample size indicated caution when interpreting the results in spite of the relatively high reliability and validity coefficients. Neverthe-

less, it was the author's opinion that Cloze Format could be used with confidence in other situations.

2. The procedures used to determine the format might well have caused the rejection of potentially reliable factors such as higher deletion rates or shorter test lengths. On the other hand, while a number of differences between correlations could not be proven statistically significant, other designs or larger samples might have provided different results.

3. The use of the teacher's judgement and the Gates-MacGinitie Reading Tests (1965) as measures on which to base claims of concurrent validity could be questioned.

4. Cloze Format was designed to be a simple placement measure within the financial and technical capability of every classroom teacher. It was not intended to be a highly sophisticated diagnostic tool. This does not mean that cloze procedure might not be used in this way either now or in the future.

5. The present study really tested only the reliability of Cloze Format as a testing instrument. While the concurrent validity found indicated that Cloze Format would be a valid reading placement device, such validity would have to be proven by usage.

Conclusions

Recognizing that the above limitations might influence later findings related to Cloze Format, both hypotheses were accepted.

Implications

1. The main implication of this study is that a reliable cloze test can be constructed to measure reading ability at the fourth year level by using the Cloze Format developed by this study. It might also be found that this format can be used in two related situations. First, using materials rated at a fourth year readability level, exceptional students in lower or higher grades might be tested using cloze tests based on this format. Second, this format might be generalized to other grade levels especially grades three to five.

2. It must be noted that the use of this format can answer only one of the questions associated with reading ability and readability. That is, once a teacher has found the cloze test scores for the pupils in his/her class, these scores must still be interpreted to determine whether the material tested is too difficult, too easy, or at a suitable level. A conversion table to suit this purpose is contained in Appendix B. This table is based on material summarized by Mork (1971) and presented earlier in Table VII.

3. The results of the present study with regard to cloze test length should caution users of cloze procedure to test the reliability of results from cloze tests containing less than 200 words or 40 blanks.

Recommendations

1. Teachers should be encouraged to use cloze procedure as one means of determining the reading abilities of their pupils and the readability of various print materials. Such a use should provide teachers with a means of making a good match of the reading abilities of their pupils with suitable reading materials. The Cloze Format developed by this study should provide a starting point for those who wish to use cloze procedure.

2. Teachers who do use cloze procedure should be encouraged to test the formats they have developed to determine whether they are reliable.

3. Reliable cloze formats should be developed for other grade levels or particular class situations. It is quite possible that other deletion rates and/or test lengths would still have adequate reliability.

4. While the decisions relating to the development of Cloze Format often were based on differences which could not be shown to be significant, some doubts were raised regarding the reliability of very short tests and low deletion rates.

5. Experimentation should be carried out to determine whether or not cloze procedure has construct validity.

6. Further experiments should be carried out to validate or modify the conversion table in Appendix B in connection with a standardized cloze test format.

7. The development of an oral cloze format should be attempted. Such a cloze test might prove very useful in determining a child's "reading capacity [Zintz, 1970, p. 55]."

8. Exploration of cloze procedure as a teaching device might lead to the development of instructional procedures which would help improve the teaching of reading.

Concluding Statement

This study developed a simple cloze test format which was shown to have split-half and alternate form reliability and to have concurrent validity. The format was designed to be used in conjunction with a device, such as a table, to relate the cloze test scores found to the readability level of the material tested. Using cloze procedure and cloze test scores in this manner, teachers have the use of a relatively simple placement device.

REFERENCES

American Psychological Association. Publication manual (1967 Revision). Washington, D.C.: APA, 1967.

American Psychological Association. Standards for educational and psychological tests and manuals. Washington, D. C.: APA, 1966.

Anderson, J. Research in readability for the classroom teacher. Journal of Reading, 1966, 9, 402-405.

Anderson, J. A technique for measuring reading comprehension and readability. English Language Teaching, 1971, 25, 178-182.

Betts, E. R. Adjusting instruction to individual needs. In N. B. Henry (ed.), Reading in the Elementary School, Chicago: University of Chicago, 1949, 48, 277.

Bloomer, R. H. The cloze procedure as a remedial reading exercise. Journal of Developmental Reading, 1962, 5, 173-181.

Blumenfeld, J. P., & Miller, G. R. Improving reading through teaching grammatical constraint. Elementary English, 1966, 43, 752-755.

Bormuth, J. R. Experimental applications of cloze tests. In J. A. Figurel (ed.), Improvement of reading through classroom practice. Newark, Del.: IRA Conference Proceedings, 1964, 9, 303-306.

Bormuth, J. R. Optimum sample size and cloze test length in readability measurement. Journal of Educational Measurement, 1965, 2, 111-116.

Bormuth, J. R. Factor validity of cloze tests as measures of reading comprehension ability. Reading Research Quarterly, 1966, 4, 358-365.

Bormuth, J. R. Comparable cloze and multiple choice comprehension test scores. Journal of Reading, 1967, 1967, 10 (5), 291-299. (a)

Bormuth, J. R. New data on readability. In J. A. Figurel (ed.), Forging ahead in reading. Newark, Del.: IRA Conference Proceedings, 1967, 12 (1), 488-492. (b)

Bormuth, J. R. The cloze readability procedure. Elementary English, 1968, 45, 429-436. (a)

Bormuth, J. R. Cloze test readability: Criterion reference scores. Journal of Educational Measurement, 1968, 5 (3), 189-196. (b)

Bormuth, J. R. Empirical determination of the instructional reading level. In J. A. Figurel (ed.), Reading and realism. Newark, Del.: IRA Conference Proceedings, 1969, 13 (1), 716-721.

British Columbia Teachers' Federation. Minutes of the 58th Annual General Meeting. Vancouver: BCTF, 1974.

Buros, O. K. (ed.). The seventh mental measurements yearbook. Vol. II. Highland Park, N.J.: Gryphon Press, 1972.

Burron, A., & Claybaugh, A. L. Basic concepts in reading instruction. Columbus: Merrill, 1972.

Byrne, M. A., Feldhusen, J. F., & Kane, R. B. The relationships among two cloze measures and divergent thinking abilities. Reading Research Quarterly, 1971, 6, 378-393.

Cattell, R. B., & Cattell, A. K. S. The culture fair intelligence test. Champaign, Ill.: Institute for Personality and Ability Testing, 1959.

Causey, O. S. (ed.). The reading teacher's reader. New York: Ronald Press, 1958.

Coleman, E. B. Experimental studies of readability. Elementary English, 1968, 45, 166-178.

Coleman, E. B., & Miller, G. R. A set of thirty-six prose passages calibrated for complexity. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 851-854.

Cronbach, L. J. Validity. In C. W. Harris (ed.), Encyclopedia of educational research (3rd ed.). New York: Macmillan, 1960, pp. 1551-1555.

Culhane, J. W. Cloze procedures and comprehension. Reading Teacher, 1970, 23, 410-413 ff.

Dale, E., & Chall, J. S. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 11-28.

Deschant, E. V. Improving the teaching of reading. (2nd ed.). Englewood Cliffs: Prentice-Hall, 1970.

Dickens, M., & Williams. F. An experimental application of cloze procedure and attitude measures to listening comprehension. Speech Monographs, 1964, 31, 103-108.

Feldman, M. E. Learning by programmed and text format at three levels of difficulty. Journal of Educational Psychology, 1965, 56 (3), 133-139.

Ferguson, G. A. Statistical analysis in psychology and education (2nd ed.). Toronto: McGraw-Hill, 1966.

Flesch, R. A new readability yardstick. Journal of Applied Psychology, 1948, 32 (3), 221-233.

Fries, C. C. Linguistics and reading. New York: Holt, Rinehart & Winston, 1963.

Galloway, P. How secondary students and teachers read textbooks. Journal of Reading, 1973, 17, 216-219.

Gates, A. I., & MacGinitie, W. H. Gates-MacGinitie reading tests. New York: Teachers College Press, 1965.

Groff, P. (ed.). Research critiques. Elementary English, 1971, 48, 675-681.

Guthrie, J. T. Learnability vs. readability of texts. The Journal of Educational Research, 1972, 65 (2), 273-280.

Hafner, L. E. Relationships of various measures to the "cloze." New concepts in college-adult reading. (13th yearbook of the National Reading Conference.) Milwaukee: NRC, 1964, pp. 135-145.

Hafner, L. E. Implications of cloze. The philosophical and sociological bases of reading. (14th yearbook of the National Reading Conference.) Milwaukee: NRC, 1965, pp. 151-158.

Hafner, L. E. Cloze procedure. Journal of Reading, 1966, 9, 415-419.

Jones, M. B., & Pikulski, E. C. Cloze for the classroom. Journal of Reading, 1974, 17 (6), 432-438.

Kirby, C. L. Using the cloze procedure as a testing technique. Use of Tests in Evaluation. Newark, Del.: IRA Conference Proceedings, 1970, 13 (4), 68-77.

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

Klare, G. R. Comments on Bormuth's readability: A new approach. Reading Research Quarterly, 1966 (summer), 119-125.

Lee, W. D. What does research in readability tell the classroom teacher? Journal of Reading, 1964, 8, 141-144.

Levine, H. F. Linguistic and paralinguistic changes in Spanish speakers learning English. English Language Teaching, 1971, 25 (3), 288-295.

Lorge, I. Readability formula--an evaluation. Elementary English, 1949, 26, 86-95.

Louthan, V. Some systematic grammatical deletions and their effects on reading comprehension. English Journal, 1965, 54, 295-299.

Manis, M., & Dawes, R. M. Cloze scores as a function of attitude. Psychological Reports, 1961, 9, 79-84.

Marshall, W. A. Contextual constraint on deaf and hearing children. American Annals of the Deaf, 1970, 115 (7), 682-689.

Moores, D. F. An investigation of the psycholinguistic functioning of deaf adolescents. Exceptional Child, 1970, 36, 645-652.

Mork, T. A. Clozing the placement gap: A new tool for administrators and teachers. Educational Leadership, 1971, 28, 763-767.

Moskowitz, S. Building your own reading tests. Instructor, 1971 (March), 52.

Oller, J. W. Scoring methods and difficulty levels for cloze tests of proficiency in English as a second language. Modern Language Journal, 1972, 56, 151-158.

Oller, J. W., & Conrad, C. A. The cloze technique and ESL proficiency. Language Learning, 1971, 21 (2), 183-194.

Peisach, E. C. Comprehension of teacher and peer speech. Child Development, 1965, 36, 467-480.

Popham, W. J. Educational statistics. New York: Harper & Row, 1967.

Preston, R. C. Ability of students to identify correct responses before reading. Journal of Educational Research, 1964, 58 (4), 181-183.

Rankin, E. F. Close procedure--its validity and utility. Starting and improving college reading programs. (8th yearbook of the National Reading Conference.) Milwaukee: NRC, 1959, pp. 131-144. (a)

Rankin, E. F. Uses of the cloze procedure in the reading clinic. In J. A. Figurel (ed.), Reading in a changing society. New York: Scholastic Magazines, 1959, 4, 228-232. (b)

Rankin, E. F. The cloze procedure--a survey of research. The philosophical and sociological bases of reading. (14th yearbook of the National Reading Conference.) Milwaukee: NRC, 1965, pp. 133-150.

Rankin, E. F. An interview with Earl Rankin. Journal of Reading, 1969, 12 (7), 574-580.

Rankin, E. F., & Culhane, J. W. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1969, 13, 193-198.

Ransom, P. E. Determining reading levels of elementary school children by cloze testing. In J. A. Figurel (ed.), Forging ahead in reading. Newark, Del.: IRA Conference Proceedings, 1967, 12 (1), 477-482.

Robinson, R. D. An introduction to the cloze procedure. Newark, Del.: International Reading Association, 1972.

Rothkompf, E. Z. Textual constraint as a function of repeated instruction. Journal of Educational Psychology, 1968, 59 (1), 20-25.

Science Research Associates (Canada) Inc. Reading laboratory series. Don Mills: SRA, 1969.

Semmel, M. I., Barritt, L. S., & Bennett, S. W. Performance of EMR and non-retarded children on a modified cloze task. American Journal of Mental Deficiency, 1970, 4, 681-688.

Skyer, G. A comparison of the cloze procedure and the short answer achievement test in determining subject matter comprehension. Graduate Research in Education and Related Disciplines, 1969, 2 (Fall), 29-45.

Slade, K., Conry, R., & Rodgers, D. Pupils' speech abilities vary by economic status. The Vancouver Province, April 11, 1974, p. 27.

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

Taylor, W. L. Recent developments in the use of "cloze procedure." Journalism Quarterly, 1956, 33, 42-48, 99.

Taylor, W. L. Cloze readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology, 1957, 41, 19-26.

Tuinman, J. J. The removal of information procedure--a first analysis. Journal of Reading Behaviour, 1971, 3 (2), 44-50.

Veatch, J. Individualizing your reading program. New York: Putnam's Sons, 1959.

Weaver, W. W. Theoretical aspects of cloze procedure. Philosophical and sociological bases of reading. (14th yearbook of the National Reading Conference.) Milwaukee: NRC, 1965, pp. 115-132.

Weaver, W. W., & Kingston, A. J. Factor analysis of the cloze procedure and other measures of reading and language ability. Journal of Communication, 1963, 13, 252-261.

Weaver, W. W., & Kingston, A. J. Recent developments in readability appraisal. Journal of Reading, 1967, 11, 44-47.

Zintz, M. The reading process. Iowa: Brown, 1970.

APPENDIX A

THE CLOZE TESTS

SAMPLE TEST 1

Pooh always liked a ___LITTLE___ something at eleven o'clock ___IN___ the morning, and he ___WAS___ very glad to see ___RABBIT___ getting out the plates ___AND___ mugs; and when Rabbit ___SAID___, "Honey or condensed milk ___WITH___ your bread?" he was ___SO___ excited that he said, "___BOTH___," and then, so as ___NOT___ to seem greedy, he ___ADDED___, "But don't bother about ___THE___ bread, please." And for ___A___ long time after that ___HE___ said nothing ... until at ___LAST___, humming to himself in ___A___ rather sticky voice, he ___GOT___ up, shook Rabbit lovingly ___BY___ the paw, and said ___THAT___ he must be going ___ON___.

"Must you?" said Rabbit ___POLITELY___.

"Well," said Pooh, "I ___COULD___ stay a little longer ___IF___ it, if you," and ___HE___ tried very hard to ___LOOK___ in the direction of ___THE___ larder.

"As a matter ___OF___ fact," said Rabbit, "I ___WAS___ going out myself directly."

"___OH___, well, then, I'll be ___GOING___ on. Good-bye."

"Well, good-bye, ___IF___ you're sure you won't ___HAVE___ any more."

"Is there ___ANY___ more?" asked Pooh quickly.

___RABBIT___ took the covers off ___THE___ dishes, and said, "No, ___THERE___ isn't."

SAMPLE TEST 2

The spear shot into ___THE___ air. Chicka let out ___A___ yell. "It works! It ___WORKS___!"

Chicka, the cave boy, ___HAD___ made the first bow ___AND___ arrow.

The first thing ___HE___ wanted to do was ___TELL___ everyone about it. But ___HE___ remembered that someone had ___CALLED___ him, A-boy-who-___TALKS___-like-a-man. They ___WOULD___ laugh at him. He ___THOUGHT___ for a moment. There ___WAS___ always his grandfather. What ___WOULD___ Ol say?

He set ___THE___ bow and arrow down ___IN___ some bushes and raced ___BACK___ to the cave. Teesa ___AND___ Mea were there.

"Where ___IS___ Ol?" Chicka panted.

Teesa ___LOOKED___ wonderingly at her grandson. "___HE___ is on guard near ___THE___ ladder."

Chicka found Ol ___STANDING___ watch near the ladder.

"___THE___ Cave Bear again," said ___OL___. "He has come closer ___TO___ the village. We have ___PUT___ more men on guard. ___THE___ beast came as far ___AS___ the ladder last night."

"___MY___ zoomer," Chicka exclaimed, "it ___WORKS___! You must see it ___RIGHT___ away."

Form A (page 1)

He remembered spring with ___ITS___ growing plants and gentle ___RAINS___, summer with its flowers ___AND___ bird songs; and he ___THOUGHT___ what happy times they ___HAD___ been for the creatures ___AT___ the Pool.

But now ___THAT___ fall was here and ___SO___ many of his friends ___HAD___ gone away, Wagtail had ___A___ growing feeling of loneliness ___WHICH___ he could not shake ___OFF___.

"I wish I could ___UNDERSTAND___," he thought one afternoon ___AS___ he swam about in ___THE___ weeds, "why everyone is ___SO___ anxious to get away ___FROM___ the winter. Will there ___BE___ no one at all ___BUT___ us Frogs left at ___THE___ Pool?"

The snapping of ___A___ branch along the water's ___EDGE___ caught his ear, and ___LOOKING___ over, he saw a ___FAT___ groundhog waddle down and ___SEAT___ himself upon a pile ___OF___ earth in the sun.

___WAGTAIL___ had never cared much ___FOR___ these queer, awkward creatures ___WITH___ their furry coats; but ___NOW___, all at once, his ___HEART___ warmed to the great ___AWKWARD___ fellow who had not ___GONE___ away with the others.

___CLIMBING___ onto the bank, he ___HOPPED___

Form A (page 2)

over, and spoke to ___THE___ groundhog. "You are not ___GOING___ away from the Pool ___ARE___ you, groundhog?" he asked ___ANXIOUSLY___, "You will stay right ___HERE___--winter or no winter, ___WON'T___ you?"

The groundhog did ___NOT___ want to be disturbed ___JUST___ then, and at the ___SOUND___ of the young frog's ___VOICE___, he moved as though ___TO___ run away, but something ___MADE___ him turn back. "Of ___COURSE___ I'm not going away," ___HE___ said. "Why should I?"

"___I___ don't know," answered Wagtail, "___ONLY___ so many creatures have ___GONE___.

Let them go if ___THEY___ want to," said the ___GROUNDHOG___. "The Blue Pool is ___A___ good enough place for ___ME___!"

"And for me," said ___WAGTAIL___, "but the robins say ___IT___ is an unhappy place ___IN___ winter."

"The robins!" growled ___THE___ groundhog.

Form B (page 1)

He remembered spring with its growing plants ___AND___ gentle rains, summer with its flowers and ___BIRD___ songs; and he thought what happy times ___THEY___ had been for the creatures at the ___POOL___.

But now that fall was here and ___SO___ many of his friends had gone away, ___WAGTAIL___ had a growing feeling of loneliness which ___HE___ could not shake off.

"I wish I ___COULD___ understand," he thought one afternoon as he ___SWAM___ about in the weeds, "why everyone is ___SO___ anxious to get away from the winter. ___WILL___ there be no one at all but ___US___ Frogs left at the Pool?"

The snapping ___OF___ a branch along the water's edge caught ___HIS___ ear, and looking over, he saw a ___FAT___ groundhog waddle down and seat himself upon ___A___ pile of earth in the sun.

Wagtail ___HAD___ never cared much for these queer, awkward ___CREATURES___ with their furry coats; but now, all ___AT___ once, his heart warmed to the great ___AWKWARD___ fellow who had not gone away with ___THE___ others.

Climbing onto the bank, he hopped ___OVER___, and spoke to the groundhog. "You are ___NOT___ going away

Form B (page 2)

from the Pool are you, GROUNDHOG ?" he asked anxiously. "You will stay right HERE --winter or no winter, won't you?"

The GROUNDHOG did not want to be disturbed just THEN , and at the sound of the young FROG'S voice, he moved as though to run AWAY , but something made him turn back. "Of COURSE I'm not going away," he said. "Why SHOULD I?"

"I don't know," answered Wagtail, "only SO many creatures have gone."

"Let them go IF they want to," said the groundhog. "The BLUE Pool is a good enough place for ME !"

"And for me," said Wagtail, "but the ROBINS say it is an unhappy place in WINTER ."

"The robins!" growled the groundhog.

Form C (page 1)

He remembered spring with its growing plants and gentle rains, SUMMER with its flowers and bird songs; and he thought what HAPPY times they had been for the creatures at the Pool.

BUT now that fall was here and so many of his FRIENDS had gone away, Wagtail had a growing feeling of loneliness WHICH he could not shake off.

"I wish I could understand," HE thought one afternoon as he swam about in the weeds, "WHY everyone is so anxious to get away from the winter. WILL there be no one at all but us Frogs left AT the Pool?"

The snapping of a branch along the water's EDGE caught his ear, and looking over, he saw a fat GROUNDHOG waddle down and seat himself upon a pile of earth IN the sun.

Wagtail had never cared much for these queer, AWKWARD creatures with their furry coats; but now, all at once, HIS heart warmed to the great awkward fellow who had not GONE away with the others.

Climbing onto the bank, he hopped OVER, and spoke to the groundhog. "You are not going away FROM the Pool are you, groundhog?" he asked anxiously. "You will STAY right here--winter or no winter, won't you?"

Form C (page 2)

The groundhog ___DID___ not want to be disturbed just then, and at the ___SOUND___ of the young frog's voice, he moved as though to ___RUN___ away, but something made him turn back. "Of course I'm ___NOT___ going away," he said. "Why should I?"

"I don't know," ___ANSWERED___ Wagtail, "only so many creatures have gone."

"Let them go ___IF___ they want to," said the groundhog. "The Blue Pool is ___A___ good enough place for me!"

"And for me," said Wagtail, "___BUT___ the robins say it is an unhappy place in winter."

"___THE___ robins!" growled the groundhog.

Form D (page 1)

"I tried to run FROM it, but the snow WAS so deep that I COULD only move slowly; and IT kept right on following-- FOLLOWING --"

"What was it?" cried WAGTAIL unable to bear the SUSPENSE any longer.

The groundhog LEANED down toward the excited LITTLE fellow and said in A serious voice, "It--was-- ONLY --my--shadow!"

"Oh!" said WAGTAIL, drawing a long breath. "THAT must have made you LAUGH, to think you had BEEN trying to run away FROM your shadow."

The groundhog HALF closed his eyes. "No," HE said quietly, "my head WAS filled with queer thoughts THAT day, and I did NOT laugh."

"But you did GET back to your burrow," COMFORTED Wagtail, "you did get BACK to your warm, comfortable BURROW, didn't you?"

"Yes," said THE groundhog, "I got back SAFELY and went to sleep AGAIN, but I have never FORGOTTEN --"

Suddenly he sat up STRAIGHT. His little black paws WERE held out in front OF him.

Form D (page 2)

"Listen!" he said SHARPLY. "Listen! It's the voice!" DON'T you hear it back AT the edge of the WOOD? Right from the doorway OF my warm burrow it IS calling!"

Wagtail listened, straining HIS ears.

A sharp north WIND blew over the meadow BRINGING down leaves and bending LOW before it the yellow GRASS.

On it came, making WAVES on the water of THE Pool and whistling through THE cattails in the marsh.

ALL at once Wagtail, too, SAT up straight.

"I think I hear something," he said. "YES, I surely hear something, GROUNDHOG, but it is not COMING from the wood. The VOICE I hear is coming FROM the bottom of the POOL.

"We must be going NOW," said the groundhog quietly.

"YES," replied the little frog.

AND, without another word, they MOVED away; one going toward A burrow in the meadow and the other toward the friendly water of the Blue Pool.

Form E (page 1)

"I tried to run from it, but THE snow was so deep that I could ONLY move slowly; and it kept right on FOLLOWING --following--"

"What was it?" cried Wagtail unable TO bear the suspense any longer.

The groundhog LEANED down toward the excited little fellow and SAID in a serious voice, "It--was--only--MY --shadow!"

"Oh!" said Wagtail, drawing a long BREATH. "That must have made you laugh, to THINK you had been trying to run away FROM your shadow."

The groundhog half closed his EYES. "No," he said quietly, "my head was FILLED with queer thoughts that day, and I DID not laugh."

"But you did get BACK to your warm, comfortable burrow, didn't you?"

"YES," said the groundhog, "I got back safely AND went to sleep again, but I have NEVER forgotten--"

Suddenly he sat up straight. His LITTLE black paws were held out in front OF him.

"Listen!" he said sharply. "Listen! It's THE voice! Don't you hear it back at THE edge of the wood? Right from the DOORWAY of my warm burrow it is calling!"

Form E (page 2)

WAGTAIL listened, straining his ears.

A sharp north WIND blew over the meadow bringing down leaves AND bending low before it the yellow grass.

ON it came, making waves on the water OF the Pool and whistling through the cattails IN the marsh.

All at once Wagtail, too, SAT up straight.

"I think I hear something," HE said. "Yes, I surely hear something, groundhog, BUT it is not coming from the wood. THE voice I hear is coming from the BOTTOM of the Pool."

"We must be going NOW," said the groundhog quietly.

"Yes," replied the LITTLE frog.

And without another word, they moved AWAY; one going toward a burrow in the meadow and the other toward the friendly water of the Blue Pool.

Form F (page 1)

"I tried to run from it, but the snow was SO deep that I could only move slowly; and it kept RIGHT on following--following--"

"What was it?" cried Wagtail unable to BEAR the suspense any longer.

The groundhog leaned down toward the EXCITED little fellow and said in a serious voice, "It--was-- ONLY --my--shadow!"

"Oh!" said Wagtail, drawing a long breath. "That MUST have made you laugh, to think you had been trying TO run away from your shadow."

The groundhog half closed his EYES. "No," he said quietly, "my head was filled with queer THOUGHTS that day, and I did not laugh."

"But you did GET back to your burrow," comforted Wagtail, "you did get back TO your warm, comfortable burrow, didn't you?"

"Yes," said the groundhog, "I got back safely and went to sleep again, but I HAVE never forgotten--"

Suddenly he sat up straight. His little black PAWS were held out in front of him.

"Listen!" he said SHARPLY. "Listen! It's the voice! Don't you hear it back at THE edge of the wood? Right from the doorway of my WARM burrow

Form F (page 2)

it is calling!"

Wagtail listened, straining his ears.

A SHARP north wind blew over the meadow bringing down leaves and BENDING low before it the yellow grass.

On it came, making WAVES on the water of the Pool and whistling through the CATTAILS in the marsh.

All at once Wagtail, too, sat up STRAIGHT.

"I think I hear something," he said. "Yes, I surely HEAR something, groundhog, but it is not coming from the wood. THE voice I hear is coming from the bottom of the POOL.

"We must be going now," said the groundhog quietly.

"Yes," REPLIED the little frog.

And, without another word, they moved away; ONE going toward a burrow in the meadow and the other toward the friendly water of the Blue Pool.

Form G (page 1)

Silently, Big Joe went ___UP___ the lake in his ___CANOE___ to a place where ___HE___ had seen moose the ___DAY___ before. He needed moose ___MEAT___ to feed his family ___DURING___ the long winter ahead. ___QUICKLY___ he pulled his canoe ___UP___ to a safe place ___ON___ the shore and started ___OFF___ into the woods with ___HIS___ gun over his shoulder.

___SUDDENLY___, the loud aroo-oo-h of ___A___ pack of wolves broke ___THE___ stillness of the woods. ___THIS___ continued until, just as ___SUDDENLY___, the smell of the ___INDIAN___ reached the nose of ___THE___ leading wolf. Instantly, they ___ALL___ disappeared--except the smallest ___WOLF___.

Big Joe caught sight ___OF___ the wolf cub near ___THE___ tree. The Indian reached ___DOWN___ quickly and grabbed the ___CUB___ by the back of ___THE___ neck. Big Joe tied ___THE___ cub's feet together and ___PUT___ him inside his coat.

___THE___ scared cub opened his ___MOUTH___ every few minutes and ___CRIED___ "Aroo-oo-h." Finally he was ___QUIET___, but by then all ___THE___ moose nearby had been ___SCARED___ away. Big Joe walked ___BACK___ to the lake, put ___THE___ cub in his canoe, ___AND___ started back to the ___CABIN___.

Form G (page 2)

When Big Joe arrived ___AT___ the cabin, he called ___TO___ his two boys, Henri ___AND___ Rene. The two boys ran ___EXCITEDLY___ down to the canoe ___TO___ see this wolf cub. "___WHAT___ will we call him? ___WHERE___ will he sleep?" cried ___THE___ boys dancing up and ___DOWN___ in excitement.

"We will ___CALL___ him Little Chuck. He ___CAN___ sleep in the corner ___ON___ some pine branches."

Rene ___AND___ Henri cut some soft branches and put them in the corner of their log cabin.

Form H (page 1)

Little Chuck grew fast. HE still wasn't very friendly. BUT one day when Rene WAS gathering birch bark and KINDLING wood for his mother, HE threw a piece of WOOD where Little Chuck could SEE it. Little Chuck ran AFTER it, picked it up IN his mouth, and took IT into his house.

After RENE took the bark and KINDLING wood into the cabin, HE picked up other sticks AND threw them where Little CHUCK could reach them. Little CHUCK had a lot of FUN running after them and TAKING them into his house. SOON there was a big PILE of sticks. Little Rene GATHERED them up and took THEM into the cabin. Gathering WOOD was fun for both LITTLE Chuck and Rene. Little CHUCK soon learned to bring BACK many things the Indian BOYS threw.

One day Rene AND Henri were playing with THE sled their father had MADE for them. They decided TO fasten Chuck's chain to THE front of it. Much TO their delight, Chuck not ONLY pulled the sled but SOON learned to stop and GO on

Form H (page 2)

command, just like ___THE___ sled dogs. Little Chuck ___SEEMED___ to like pulling the ___SLED___ as much as Rene ___AND___ Henri liked to ride.

___ONE___ day Little Chuck pulled ___RENE___ and Henri on their ___SLED___ as far as the ___LITTLE___ village on the lake. ___THAT___ night he was very ___TIRED___. After he ate his ___DINNER___, he curled up on ___HIS___ pine branches to go ___TO___ sleep. Henri brought him ___A___ pan of fresh spring water, and as he put it inside Little Chuck's house, he reached over and patted Chuck's head.

Form I (page 1)

One day the boy's MOTHER said to Big Joe, "WE need flour and salt."

BIG Joe answered, "I will TAKE the dog sled and GO to the village tomorrow."

EARLY the next morning, he WENT out to harness the SLED dogs. He found Tito, THE youngest, was sick. The OTHER sled dog, Toro, could NOT pull the sled alone, ESPECIALLY when it was loaded.

RENE and Henri said, "Try LITTLE Chuck. He knows how TO pull a sled. He IS a strong wolf."

Big JOE had seen how well LITTLE Chuck could pull the BOYS' sled, so he said, "ALL right, we will try!"

THE Indian boys were delighted WHEN their father started off WITH the sled pulled by TORO and Little Chuck. Big JOE ran behind the sled MOST of the time, but HE rode on it when THEY were going downhill or WHEN the trail was smooth.

BEFORE noon, he reached the VILLAGE and went to the TRADING post. There he traded HIS beaver skins for flour AND salt. After he had EATEN his lunch and fed THE

Form I (page 2)

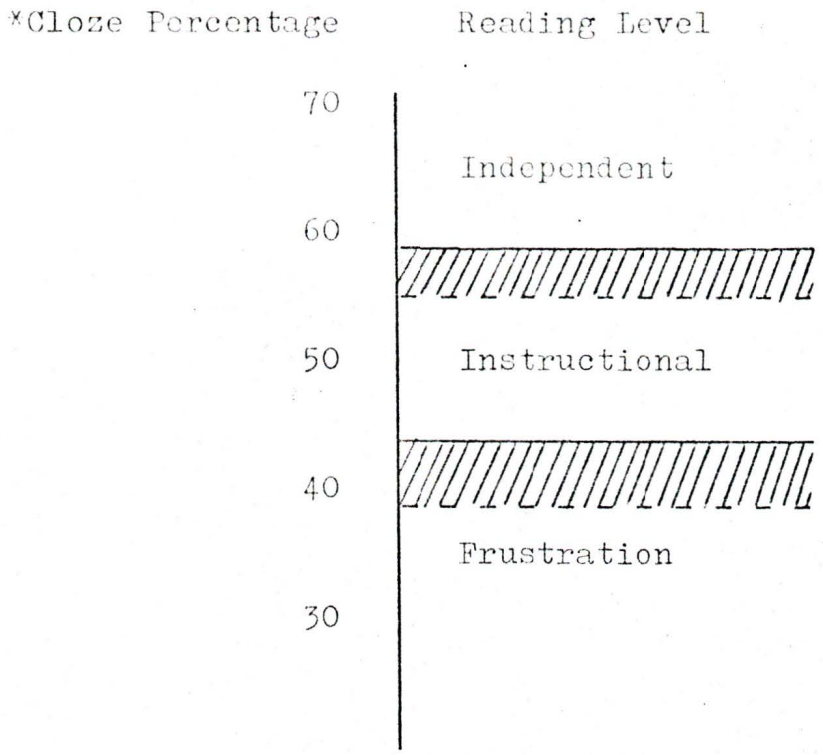
animals, he loaded the GROCERIES on the sled and STARTED for home.

Not long AFTER he left the village, THE weather changed, and it BEGAN to snow. Big Joe HAD trouble seeing the trail AS it snowed harder. When THEY were still about five MILES from home, the snow GREW very deep. Old Toro WAS so tired he could HARDLY keep up with Little CHUCK.

As the snow kept GETTING deeper, Big Joe had trouble keeping up with the sled.

APPENDIX B

Cloze Percentages and Their Reading Level Equivalents



*Using the format developed in this study, to find the cloze percentage multiply by 2.

Solid lines indicate the divisions outlined by Mork (1971).

Shaded areas indicate the possible lower limits of the instructional and independent levels.

VITA

Surname: WAIMSLEY Given Names: DONALD PERCY

Place of Birth: NEW WESTMINSTER, B.C.

Date of Birth: MARCH 1, 1941

Educational Institutions Attended,
with Dates of Entering and Leaving:

VICTORIA COLLEGE	1959	to	1962
UNIVERSITY OF VICTORIA (part-time)	1963	to	1968
UNIVERSITY OF VICTORIA (part-time)	1968	to	1974
	____	to	____

Degrees, Diplomas, Etc., Awarded,
with Dates and Names of Institutions:

B.Ed.	1968	UNIVERSITY OF VICTORIA

Honours and Awards:

Publications:

PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis or dissertation (the title of which is shown below) to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis/Dissertation

CLOZE FORMAT FOR FOURTH YEAR PUPILS

Author: [redacted]

Signature

DONALD PERCY WALMSLEY

Name

Date