THE VALIDITY AND APPLICABILITY OF TWO MODIFIED CLOZE

PROCEDURES (BEGINNING OF THE PAGE PROCEDURE AND

"INSTANT" BEGINNING OF THE PAGE PROCEDURE) MEASURED

AGAINST THE STANFORD DIAGNOSTIC READING TEST AND EQUATED

WITH THE CLOZE PROCEDURE AND FRY READABILITY GRAPH.

By

DIANNE PARKINSON

B.A., Simon Fraser University, 1971

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Department of Education(Reading)

We accept this thesis as conforming to the

required standard

THE UNIVERSITY OF BRITISH COLUMBIA

© October, 1980

Dianne Parkinson, 1980

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of _Education_

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date _May 2/81_

The validity and applicability of two modified cloze
procedures (Beginning of the Page Procedure and the
"instant" Beginning of the Page Procedure) measured
against the Stanford Diagnostic Reading Test and
equated with the cloze procedure and the Fry Graph.

ABSTRACT

This correlational study examined the Beginning of
the Page Procedure (B.O.P.P.) and the "instant"
Beginning of the Page Procedure as measures for
assessing readability. One hundred ninety-six grade
nine students (106 male and 90 female) took part in the
study and their scores on the cloze procedure, the
B.O.P.P. and the "instant" B.O.P.P. were correlated with
the Stanford Diagnostic Reading Test Form A - Blue Level
(hereafter referred to as the Stanford Diagnostic).
The Stanford Diagnostic was used as the anchor test and
the students were randomly assigned to each of the three
groups. Analyses included the calculation of means
associated with the Stanford Diagnostic scores for each
subgroup, and analysis of the variance between sexes
within each subgroup. An equivalency table is provided
which estimates the Stanford Diagnostic scores for a
given cloze procedure, B.O.P.P. or "instant"
B.O.P.P. score. Using the Stanford Diagnostic grade

score equivalent to 40 percent on the cloze procedure, the readability level of the passage was determined. This was then compared to the readability level estimated by the Fry Graph.

Respective correlations of .53 and .67 were found between the B.O.P.P. and "instant" B.O.P.P. with the Stanford Diagnostic suggesting both are good indicators of the students' ability to handle the given passage. Similarly the Fry Graph and the Stanford Diagnostic, grade score equal to 40 percent on the cloze procedure, found the passages to be at virtually the same level of difficulty. All results, however, were limited to the passage studied and should not be generalized to other materials.

When a significance level of .05 was used no significant difference was found between the male and female performance levels on any of the tests administered .

Table of Contents

Chapter III

A DESCRIPTION OF THE STUDY

Chapter IV

ANALYSIS OF DATA, SUMMARY, CONCLUSIONS AND

IMPLICATIONS, RECOMMENDATIONS FOR FURTHER STUDY

List of Tables:

List of Figures:

Figure

CHAPTER I

The Problem

Rationale for the Study

One wishes that he might more frequently find
on the secondary level materials for the
various instructional units in the content
areas on levels easier and more difficult than
those commonly used for the grade level.
Materials of this type are a must if the high
school teacher is to build his program on what
he knows of the way young people grow - some
slower, others much faster than the average
for the grade. (Bormuth 1967, p.291)

Each day students face the frustration of having
materials assigned to them that they cannot comprehend.
Educators are faced with the problem of trying to
determine what materials are suitable for a particular
student. The trend has been to develop relatively
simple and fast techniques for determining the reading
level of the given material, the reading capacity of the
student and the student's ability to deal with that
material. The matching of a student's reading level
with the reading level of assigned readings is called

readability. Techniques for determining readability
have developed in three basic directions: the
readability formula, the informal inventory and the
cloze procedure. Of these the informal inventory is
likely the least used for as Bormuth (1968) suggested,
it is time consuming and requires a relatively high
degree of training on the part of the teacher. Pennock,
(1973) further cautioned that " In reading tests where
the student is asked to answer questions, his score is
influenced not only by the passages read, but also by
the quality of the questions and his comprehension of
them" (p. 37). This criticism may be leveled to some
degree at standardized tests but more importantly at
tests that have not been subjected to the rigors of
standardization.

For the average classroom teacher a readability
formula in conjunction with a standardized test or the
cloze procedure would appear to provide the most
expedient solution to the problem of providing students
with reading materials at their own reading level. The
emergence of the readability formula from its complex
and time consuming beginning, to its present quick
scoring formulas is traced in the review of the
literature, chapter II.

Although a variety of uses have been developed for
the cloze procedure, little has been done to streamline

it when used as a readability measure. Educators are
still required to select a book, type six to twelve 250
word passages, have students fill in the deleted words,
determine the mean score on each of the passages, add
all the means together and divide by the number of
passages administered. This results in a score which is
the mean of means and the passage with the mean closest
to this score is considered to be representative of the
book. Granted this process need be done only once per
book but as Pennock (1973) suggests, " ...few classroom
teachers have the time and facilities for cloze
procedure test production as a means of assessing the
difficulty of each book" (p. 38). He does suggest that
a reading coordinator might construct such tests and
also proposes an alternative to the traditional cloze
procedure sampling process. (see Chapter 2 -
Modifications on the Cloze Procedure.)

McCabe (1979) has proposed a process which could
drastically cut the time required for the whole cloze
procedure. He calls his proposal B.O.P.P. - "Beginning
of the Page Procedure". Following McCabe's
instructions;

> ...the teacher must first type an intact
> passage onto a ditto master. Second, a strip
> of paper, which is approximately six inches

long (15cm) and 1/2 inch (1.25 cm) wide, is cut from a plain piece of paper. This strip of paper is then taped to the back of the ditto master, approximately one inch (2.5 cm) from the left hand margin. The stencil is then inserted into a Ditto reproducing machine and copies of the B.O.P.P. are made". (p.199)

An interesting aspect of the B.O.P.P. is the deletion of part words as well as whole words. Readers are asked to fill in all missing words or parts of words. They may also be instructed to cross out parts of words which do not appear appropriate and replace these with something they feel is appropriate. McCabe proposed further variations on the cloze procedure by suggesting that broader strips may be used to focus the reader's attention on larger segments of information or that the strip of paper be moved to create a "Middle of the Page Procedure", M.O.P.P. or an "End of the Page Procedure," E.O.P.P .

The major thrust of McCabe's proposal appears to focus on an abbreviated cloze procedure; one that is less time consuming and therefore more likely to be used by the classroom teacher. McCabe goes so far as to suggest that an "instant B.O.P.P." could be created by placing a strip of paper (15 cm long and 2.5 cm wide) one inch from the left hand margin of any book. That

page could then be photocopied to produce an "instant" B.O.P.P., which requires no typing.

If field studies on McCabe's proposal prove the results of the B.C.P.P. and the "instant" B.O.P.P. to be equally as valid as those of the cloze procedure, every nth word deleted, then he has created a one step readability formula with all the advantages of the cloze procedure discussed in the review of the literature, Chapter II. He has also created a quick means of determining a student's ability to deal with the given reading material and a technique more readily usable by the classroom teacher.

This study investigated the validity of the B.O.P.P. and the "instant" B.O.P.P. when compared with the cloze procedure, every fifth word deleted, and the results of a standardized test. The passage used in the study was screened using the Fry Readability Graph. Subjects were a random selection of grade 9 students in one Junior High. The effectiveness of the B.O.P.P. and the "instant" B.O.P.P. were measured in terms of the correlation of their results with the comprehension section of the Stanford Diagnostic Reading Test, Blue level, Form A (hereafter, Stanford Diagnostic).

Objectives of the Study

The major objective of this study is to answer the following questions.

1) How related are the cloze procedure, the Beginning of the Page Procedure (B.O.P.P.) and the "instant" Beginning of the Page Procedure to the Stanford Diagnostic ?

2) Is the instructional level of 40 - 59 percent using the cloze procedure equivalent to the same scores on the B.O.P.P. and the "instant" B.O.P.P?

3) Does the Fry Readability Graph estimate the readability of the given passage to be the same as the Stanford Diagnostic grade equivalent for 40 percent on the cloze procedure ?

4) Are the performance levels of males and females statistically different on the Stanford Diagnostic, the cloze procedure, the B.O.P.P. or the "instant" B.O.P.P?

Definition of Terms Used

B.O.P.P.   —   Beginning of the Page Procedure.   A
procedure that instructs the teacher to type
an intact passage onto a ditto master.   Next a
strip of paper approximately 6 inches long
(15 cm) and 1/2 inch (1.25 cm) wide is
attached to the back of the ditto, one inch
(2.5 cm) from the left-hand margin.   No part
of the first or last sentence should be
covered.   The stencil is then run on a
reproducing machine to create copies of the
B.O.P.P.   The deletions in the
B.O.P.P. include whole words and part words
and the reader is instructed to fill in the
missing parts.   For the purpose of this study
only responses that exactly matched deletions
were considered correct.   It was also noted
that a 1/2 inch deletion on a page typed on an
average typewriter, was equal to 1/10th the
spaces on the typed line (see appendix H).

"Instant" B.O.P.P.   —   this procedure is like the
B.O.P.P.   but requires no typing.   The strip
of paper is placed over a page in a book or
other printed material making sure to leave

the first and last sentence intact. The page
is then photocopied to produce the desired
number of copies. For the purpose of this
study the reader's responses once again had to
exactly match the deletions to be counted as
correct. The width of the strip of paper was
considered to be equal to 1/10th the number of
spaces on a full line in the particular
material being studied and not the 1/2 inch
suggested by McCabe. This modification to
McCabe's proposal was made to take into
account the variety in size of type found in
printed materials (see appendix I).

Spaces on a line — this includes all letters on a
given line, all punctuation and all spaces
between words. A full line is one which goes
from the left hand margin to the right hand
margin.

Cloze Procedure — a passage of at least 250 words is
chosen. The first and last sentence are left
intact and every fifth word is deleted in the
remainder of the passage up to a maximum of
fifty deletions. The deletions are replaced
with blanks of standard length and the reader

is instructed to fill in the blank with the exact word that has been deleted. Only exact replications of deleted words are·scored. The total correct responses are multiplied by two to give the percentage score.

Stanford Diagnostic Reading Test, Blue Level, Form A —
is designed to· measure reading abilities of students in grades 9 through 12 and in community colleges. For the purpose of this study only that section of form A which pertained to comprehension was administered.

Fry Readability Graph —— A quick scoring readability measure that yields a reading score somewhere between a student's instructional and frustration level.

Independent reading level —— The level at which a person can read and understand material without any assistance. This is usually considered to be equivalent to a 90 percent score on a multiple-choice comprehension test based on material at the same level.

Instructional reading level —— The level at which a

person can read and understand material with the aid of an instructor. This is usually considered to be equivalent to a 75 percent score on a multiple-choice comprehension test.

Frustration reading level — The level at which a person is unable to read and get meaning from a passage even with the aid of an instructor. This is usually considered to be equal to a score of less than 50 percent on a multiple-choice test.

Research Questions

1. . Will the cloze procedure, the Beginning of the Page Procedure and the "instant" Beginning of the Page Procedure be positively correlated with the Stanford Diagnostic ?

2. Are the scores yielded by the cloze procedure, the Beginning of the Page Procedure and the "instant" Beginning of the Page Procedure equivalent?

3. What is the difference between the Fry estimate of readability for the passage and the Stanford Diagnostic grade equivalent for 40 percent on the cloze procedure?

4. Is there a significant difference between the performance levels of males and females on the Stanford Diagnostic, the cloze procedure, the B.O.P.P. or the "instant" B.O.P.P?

Basic Assumptions

    For the purpose of this study the following assumptions were made.

1. The Stanford Diagnostic, constituted a valid measure of a student's reading achievement or reading grade level when one and a half years was subtracted from the reading grade score to determine the student's instructional reading level. (Burmeister, 1974)

2. The Fry Readability Graph gave a reasonable indication of the cloze procedure passage being studied when one and a half years was added to the calculated score in order to obtain the level at which the material could be used for instruction. (see pg. 97)

3. The students' responses to the passages (including the cloze procedure, B.O.P.P. and "instant" B.O.P.P.) represented an honest attempt to replace the deleted word, words or parts of words.

4. The cloze tests selected were equal in difficulty to any other cloze tests that could have been made over the same passage.

5. The subjects selected for treatments one, two, and three, represented the same population.

Limitations of the Study

1. Only one form of each cloze test was used for the study and one cannot be sure that the cloze tests chosen were equal in difficulty to any other cloze test that could have been made over the same passage.

2. The population studied was limited to the grade nine body in one school in a suburban middle class district.

3. The three treatments were given to three different groups (assumed to be equal) and as such the effect of each treatment was not so comparable as it might have been had all three treatments been given to each subject. This research, however, relied on interrupting other teacher's classes and to avoid further interruptions, only one form of each test was administered.

4. The grade equivalent for students scoring at the top end of the Stanford Diagnostic was designated as graduate level. This did not distinguish between the different scores within this range.

Organization of the Remainder of the Study

Chapter II presents a review of the most frequently used readability formulas and the literature pertaining to these formulas. The section of this chapter dealing with readability formulas is designed only as a brief overview and as an introduction to the cloze procedure which is the major thrust of the chapter. A review of the research dealing with cloze procedure is traced from the earliest attempts to validate the cloze procedure to its present position as a valid and useful measurement for both researchers and practitioners.

Chapter III provides a detailed description of the study including: the selection of the subjects, procedures for measuring the students' reading achievement, the selection of the passage to which the cloze procedure is applied, and the construction administration and marking of the cloze procedure, the Beginning of the Page Procedure and the "instant" Beginning of the Page Procedure. The research questions and the analysis of data are also presented in this chapter.

Chapter IV presents analysis and discussion of the research questions. The conclusions and implications for future study are also included in this chapter. References and appendices are located immediately after chapter IV.

CHAPTER II

REVIEW OF THE LITERATURE

Readability Formulas

The readability formula appears to be the most
widely, if not the most wisely used technique for
assessing readability. Burmeister (1974) suggested that
the determination of readability through the use of a
readability formula was a two step process, a) requiring
that a standardized test be administered to establish
the students' reading level, and b) requiring that a
readability formula be applied to determine the level of
the given material. Ideally, students were then matched
to materials at their level.

Burmeister cautioned that silent reading tests
administered above the primary grade level tended to
yield grade scores that were equal to the students'
frustration level and that it was necessary to drop one
to two full grades from the test results in order to
determine the students' instructional level.

The need to identify reading levels for both the
students and material was emphasized in a study by
Galloway (1973) who pointed out that teachers often
choose texts based on content and judge the readability

of the text in relation to their own reading ability and their familiarity with the subject. She cited the example of one Toronto high school where all text books, except one, were found to be too difficult for non college bound students.

Teacher time is an important factor in the decision to use a formula. Klare (1963) suggested that speed of application as well as the predictive accuracy of the formula were the characteristics most frequently considered by users of formulas. To date there are well over thirty readability formulas available for use, many of which contain extended calculations which may require manual aids or even computors. As these devices are not readily available to the classroom teachers such formulas were not considered. Only those measures considered by the literature to be both quickly administered and relatively accurate were included.

Lorge Readability Formula

One of the early formulas to receive wide use was developed by Irving Lorge in 1939. Lorge was the first of many to use the McCall Crabbs Test Lessons in Reading (hereafter referred to as McCall Crabbs Test Lessons) as a criterion for his study. By correlating his formula to the McCall Crabbs passages (which had already been

graded), he was able to develop a three factor formula
which computed average sentence length, number of
prepositional phrases per 100 words and a count of the
number of hard words not on the Dale list of 769 words.
This formula gave the grade placement value of the
average reading ability required to answer 75 percent of
the test questions correctly (Klare, 1963). Some years
later the original formula was corrected and the grade
placement was changed to correspond to 50 percent
comprehension of test questions.

Dale and Chall (1948) criticized the Lorge formula
saying that the 769 easy words list did not
differentiate between the higher levels of difficulties
(see appendix A).

Flesch Readability Formula

The Flesch formula was the next major readability
measure to appear. Flesch developed a relatively simple
and accurate three factor formula which was correlated
to the McCall Crabbs Test Lessons using a 50 percent
comprehension level. Unlike Lorge, Flesch was very
skillful in gaining publicity for his formula and
brought attention to the concept of readability to most
areas of mass communication. Flesch's criticism of
earlier formulas related to what he felt was their

failure to provide a formula suitable for adult materials. He felt that they emphasized vocabulary at the expense of other factors and paid too little attention to .the role of abstract words in determining difficulty. The Flesch formula calculated sentence length, number of affixes and number of personal references (see appendix B).

Dale and Chall (1948) criticized the Flesch formula saying that people count affixes differently and therefore don't count the same number. They also felt that personal references could not be subtracted from difficulty if those references were not familiar to the reader. In that same year Flesch revised his formula having found the count of affixes too time consuming and the count of personal references misleading. The result was two new formulas still based on the 1925 McCall Crabbs Test Lessons. These were the Reading Ease formula and the Human Interest formula. Klare (1963) suggested that the Human Interest formula was not popular with users but the Reading Ease formula became widely used (see appendix C).

Dale - Chall Readability-Formula

In 1948 Dale and Chall produced their own formula which (along with the Flesch Reading Ease formula) quickly became one of the two most used formulas. The Dale-Chall formula used the McCall Crabbs Test Lessons as had the aforementioned formulas. Their aim was to attain a grade score equivalent to 50 percent comprehension on each of the passages. The formula consisted of only two factors, the average sentence length in words and the number of words outside the Dale list of 3,000.

The Dale list was constructed after testing grade four students on their knowledge of 10,000 words. Words considered known by 80 percent of the subjects were placed on the list of familiar words.

Klare (1975) reported that the Dale-Chall formula was tested against the 1925 McCall Crabbs Test Lessons and found to correlate at the 70 percent level. Dale and Chall (1948) reported that the formula was also validated on Health and Social Studies materials and a correlation of .90 - .92 was found with the judgements of expert teachers in the field and with actual reader comprehension.

In 1958 Plowers, Sumner and Klare recalculated both the Flesch and the Dale-Chall formulas based on the 1950 McCall Crabbs Test Lessons. They found that the Flesch

Reading Ease formula correlated at the .64 level with the 1950 McCall Crabbs Test Lessons while they had correlated at the .70 level with the 1925 scores. The Dale - Chall formula had a correlation of .71 with the 1950 scores which is virtually the same as the .70 correlation with the 1925 scores. As a result of the consistency in the Dale - Chall formula, Klare (1963) suggested that it was the most accurate general-purpose formula up to 1960 (see appendix D for formulas and corrected grade levels).

Fry Readability Graph

The Fry Readability Graph first appeared in 1965. Fry's Graph had two variables, syllables per 100 words and words per sentence. These two variables were entered on the graph and the readability score was then read directly from the graph. Pauk (1969) and later Vaughan (1976), in a study at the University of Arizona, found that Dale - Chall and Fry scores consistently agreed within one grade level. Klare (1975) also reported that the Fry Graph had been validated on both primary and secondary materials and the scores read from this graph had correlated highly with several well known formulas (see appendix E for graph).

SMOG Readability Formula

McLaughlin (1969) published his SMOG readability formula which he believed was simpler, quicker and more valid than earlier methods. McLaughlin stated that there was no need to count all syllables. His formula counted the number of words of three or more syllables (polysyllable count) within 30 sentences. The SMOG formula operates on the premise that a) in English longer words are usually more precise and therefore extra effort is needed to identify their full meaning, and b) longer sentences usually have a more complex grammatical structure and the reader has to retain several parts to understand the whole (McLaughlin, 1969).

The SMOG formula, like the majority of formulas considered here, was validated against the McCall Crabbs Test Lessons but instead of using the 50 - 75 percent criteria used by previous formulas, McLaughlin used the 100 percent criterion and therefore found material to be one and a half to two grades higher than other formulas. The McLaughlin formula determines the independent level of the material, whereas the other formulas determine the frustration to instructional level of the material. For example: if the Dale - Chall formula finds a book to be at the grade 4 level, using the 50 percent criterion, the book will be near the frustration level for a child reading at the grade 4 level even if he/she has teacher

assistance. The McLaughlin formula is likely to find the same book to be at the grade 6 level for only a student at that level could understand the material without teacher assistance.

The standard error on the SMOG is 1.5 grades, slightly higher than for other formulas, but McLaughlin feels that the grade level corrections made by other formulas make his comparable (see appendix F).

Bormuth Readability Formula

Bormuth, (1969b) questioned all readability formulas to that date and pointed out that no research had ever been published on the norms for the McCall Crabbs Test Lessons against which most formulas had been validated. Bormuth correlated cloze procedure percentage scores with reading achievement grade placement scores for the same students. Grade placement scores corresponding to the 35, 45, and 55 percent cloze procedure scores were determined. Using a cloze procedure criterion score of 45 percent, Bormuth found a correlation of .83 and a cross validation of .92 with the difficulty of the passage from which it was taken.

Bormuth criticized traditional means of judging the suitability of the difficulty level of materials, saying they were based upon arbitrary choice. He maintained

that his formula selected a level of performance which represented a level of comprehension difficulty at which negative outcomes of reading were minimized and positive outcomes were maximized. Bormuth's "formula" appeared to have other advantages in that it had the capacity to measure sentence difficulty or even word difficulty along with passage difficulty. However, Bormuth cautioned that his study could account for only 85 percent of the observed variables in the passages. He concluded that the test still lacked validity, that even this type of test could be fooled by easy words and difficult concepts, and that further research was needed.

Cautions Concerning Readability Formulas

Any of the five formulas discussed, excluding Bormuth's study, would appear to give the user a reasonably similar level of readability. The question is, how much credence should devices for measuring readability be given?

Readability formulas, whether they use word lists or a syllable count to measure word difficulty, are not able to take into account well known words used in a symbolic or metaphoric sense (Dale and Chall, 1948). Also, they cannot measure the author's style, the effect

of typography or format on the reader, the interest level of the material, or the readers purpose, background, and familiarity with the subject. These factors certainly affect readability but do not appear to be measured by the readability formula (Keonk, 1971 & Daines and Mason 1972). Emphasizing the need for caution, Klare (1976) pointed out that the words in a sentence or the sentences in a paragraph, could be scrambled and most formulas would find the readability level to be the same as the unscrambled version.

Klare (1976) cited McLaughlin, (1966) who suggested that when reader's background information or level of interest was high, then readability became less critical. He also stated that "Formula scores are, at best, first approximations to difficulty for readers, and human judgements are needed along with the scores" (p.141).

Klare (1976), Hansell (1976), McLaughlin (1969), and Dale and Chall (1948) cautioned that readability scores were most useful when thought of in terms of a range of difficulty rather than a precise grade level. Vaughan (1976) suggested that this range should be plus or minus one full grade.

The general-use readability formulas, therefore, provide a useful guideline for the subject teacher but they must be used in conjunction with teacher judgement.

The Cloze Procedure

The cloze procedure was first introduced by Wilson L. Taylor in 1953 and at that time was seen as a new tool for assessing readability. In the twenty-seven years since its conception, researchers have found a myriad of uses for the new technique. This study, however, focused only on cloze procedure as a measure of comprehension and readability.

Taylor 1953, explained that the term cloze was derived from a theory in gestalt psychology which suggests there is a human tendency to complete a familiar but incomplete pattern - to "see" a broken circle as a whole one, for example, by mentally closing up the gaps.

Taylor pointed out that existing readability formulas were not sensitive enough to style and he cited examples where formulas found the writings of Gertrude Stein and James Joyce to have a low readability level. He reasoned that this was because readability formulas which take into account such things as short and common words and short and simple sentences, have no means of measuring concept load. Taylor 1953 stated:

> Cloze procedure counts no such elements. It seems, however, to measure whatever effects elements actually may have on readability. And it does so at the same time it is also

taking account of the influences of many other

factors readability formulas ignore. (p.417)

This theory was supported by Russell (1978) who stated that cloze procedure had the capacity to measure such factors as sentence structure, size of print, concept load, interest, language, and even author style. In two studies in 1953, Taylor attempted to show that:

1) the cloze procedure would rank passages taken from Flesch's How to Test Readability in the same order as did the Flesch formula and the Dale-Chall formula. (Klare 1963)

2) that the cloze procedure would "handle" passages that the two standard formulas could not due to their inability to handle concept load.

For experiment 1 it was found that the cloze procedure ranked the passages in the same order as did the formulas and for experiment 2 that the cloze procedure came closer than either formula to properly ranking the readability levels of the passages (p.427).

Taylor (1953) admitted that readability formulas did have some advantages over the cloze procedure in that they were quicker and easier to apply and for "standard" materials they seemed reasonably accurate. A problem arose in that it was difficult to determine in

advance, which materials were "standard". He concluded that :

> It is a little unreasonable that a single readability score for an article on cattle breeding should apply alike to residents of Texas "cow country" and metropolitan Brooklyn. In such cases it appears that the user of a formula might employ cloze procedures to check up on his results. (p.433)

and also that:

> ...a cloze score appears to be a measure of the aggregate influences of all factors which interact to affect the degree of correspondence between the language patterns of transmitter and receiver. (p.432)

In 1957, Taylor stated that the readability technique operated on the assumption that " a) the more readable a piece of writing is, the better understood it will be even if some words are left out, and b) the better writing is understood, the more likely it is that a reader can guess what words are missing" (p.19). This was supported by Hafner (1966) who stated that the individual's choice (of words) was an index of his/her ability to comprehend reading matter.

Showing an early interest in cloze procedure, Bormuth (1966) criticized existing readability formulas

stating:

> It is problematic whether presently available
> formulas help more than they hinder. Because
> these formulas are easy and inexpensive to
> apply, they enjoy widespread use by publishers
> and educators. Publishers use them for
> "adjusting" the difficulty of instructional
> materials, and educators use them to decide if
> instructional materials are suitable for
> students at a given level of reading
> difficulty. Chall (1958) has made a strong
> case that formulas are not sufficiently
> accurate to warrant either of these uses.
> (p.81 - 82)

Bormuth (1967), pointed out that until 1967 there
was no means of determining whether a given cloze
procedure score represented an "acceptable" level of
performance by a given student. He compared cloze
procedure and multiple-choice questions produced from
the same materials and found that the correlation was
significant. Still, in 1967, he found it necessary to
caution readers that the use of the cloze procedure was
quite complex. By 1968, after further experiments, he
was able to conclude:

> a) cloze readability tests provide a valid measure
> of a student's reading comprehension ability

b) the cloze readability procedure provides a valid method of measuring the comprehension difficulties of passages

c) cloze readability scores can be used to judge the suitability of materials for a given student.

The cloze procedure became an object of interest and respect as is evidenced by the rapid upsurge of studies dealing with this topic in the 70's. The cloze procedure has been validated not only as a readability device but also as a teaching device.

Structure of Cloze Procedure Passages

The original study by Taylor, (1953) set no optimum number of words per passage nor did it specify the number of deletions per passage , rather it suggested every nth word be deleted or that random deletions be made. In a pilot for the 1953 study Taylor found that a one in five deletion system discriminated between subjects better than did a system involving fewer deletions. In a 1956 study Taylor concluded that "it appears that an every fifth-word deletion system spaces blanks as far apart as they need be" (p.45). The every fifth word deletion appears to be generally accepted in the literature (Bormuth 1968, 1968b). Rankin and Culhane (1970) suggested that the every fifth word

deletion system was suitable for narrative material but that every tenth word might be more suitable for textual fact laden material. This was supported by Potter (1968) who suggested that in some instances deletions should be one in twelve. MacGinitie (1961) reported that he found no statistically significant difference in restoring omitted words when every 24th, 12th or 6th word was omitted but he found omitting every 3rd word made restoration difficult. Oller (1975) tested every 5th, 10th, 25th plus deletions and found, contrary to MacGinitie, that the longer the surrounding context, the easier the cloze item.

Not all researchers have accepted the principle of random or every nth word deletions. Taylor (1956) experimented with easy word versus hard word deletions and Rankin (1959) experimented with structural versus lexical deletions. Both concluded that the any-word deletion system produced generally superior results. In conclusion, the literature seems to indicate that the any-word deletion system is the most practical when measuring general comprehension or readability and the every fifth word deletion system is most popular when researching narrative material. The question of the necessity for fewer deletions in fact laden materials appears to be unresolved.

Passage Length

The length of a passage required to produce a valid cloze result on the cloze procedure has been of concern to researchers. Taylor (1956) suggested a minimum 250 word passage. Bormuth (1968), Rankin (1970), and Walter (1974), concurred with this opinion and the trend in the literature appears to be to use this minimum.

Taylor (1956) suggested that cloze passages should contain fifty items which he felt was a large enough sample to allow easy and hard words to cancel out. Bormuth (1967) stated, "The test, for reasons of both convenience and reliability, should contain exactly fifty items" (p.294). There appears to be little controversy in the literature over the fifty word deletion practice and most researchers appear to adhere to it.

Boyce (1974) reported little concurrance in the literature as to amount of uninterrupted prose that should be left before deletions began. Some studies started deletions from the first sentence, some left the first sentence or two, and still others left as much as the first paragraph intact. Boyce (1974) cited Oller (1972) who wrote " as is customary, the first and last sentence of each paragraph were left intact" (p. 15). Bormuth (1969b) and Rankin and Culhane (1969) reported using this procedure but many neglected to report this

aspect of their study.

Pre-cloze Versus Post-cloze

Another aspect of the cloze procedure that is of concern to researchers is what Rankin (1965) has called pre-cloze and post-cloze -- pre-cloze being a cloze test taken before reading the original unmutilated passage and post-cloze being a test taken after reading the mutilated passage. Taylor, (1956) found post-cloze test results correlated slightly higher with scores on comprehension tests. Bormuth cited Rankin (1957) whose results supported those of Taylor. Bormuth (1968) however, theorized that these results were "...probably the result of scores being more variable than when students had not read the passage..." (p.192). He suggested that this effect could be obtained more easily by adding a few items to the test. In 1968, he reported that "research shows that the two methods are equally valid" (p. 193). Because of savings in time and preparation he felt it was more desirable to use the pre-test technique. The pre-test technique has not gone without criticism. Boyce (1974) felt that subjects who filled in blanks without an overview might treat the deletions as a series of subtests, accounting for some answers which were wrong in the total context, appearing

correct in the limited context of a sentence or group of words.

Space Length

In determining the length of spaces to be left in place of the deleted word in the mutilated passage, Taylor (1953) proposed that all spaces should be of uniform length so as to give the subjects no information on word length. This was reiterated by Taylor (1956, 1957), Bormuth (1967 1968, 1969) and Bortnick and Lopardo (1976) to name just a few. Although the use of a uniform space length appeared to be widely accepted in the literature, not all researchers agreed it was necessary. Anderson (1971) and Spooncer (1974) compared passages using the uniform space to passages using spaces the same length as the deleted word and found no significant difference between the two forms.

Selecting a Representative Passage

Much attention has been given to the mechanics of the cloze procedure but how can researchers have confidence that the passage they have chosen is representative of the material being tested? Bormuth (1968) suggested that six to twelve passages be randomly selected from the material being considered and that passages using the cloze procedure (based on a minimum of 250 words and 50 deletions) be administered to 25 to 30 students. The mean score on each test was to be calculated and then the mean of means calculated. The passage with the mean score closest to the mean of means was to be selected as the representative passage. Bormuth emphasized that the more tests made, the more representative would be the passage chosen. He also cautioned that materials that showed a great deal of variance from passage to passage would be ill suited to this technique.

Bormuth (1964) explained that within a cloze passage using every fifth word deletions, there were five possible tests and he found that there was a significant difference between the mean score on each. He did point out that the difference tended to diminish as more items were included. He concluded that using a single cloze procedure test over a passage should probably be avoided when precise determinations of

difficulty were needed and he cautioned that if one passage were used, then observed differences must be significantly different so as to assure the differences did not arise solely because of differences in test forms.

Scoring Methods

Much controversy has surrounded the question of scoring the cloze test. Does one accept synonyms or will only the exact word deleted from the passage be accepted? Despite many intuitive feelings to the contrary, the bulk of research tended to support exact word replacement. Taylor (1953), Rankin (1959) , Ruddell (1964), Bormuth (1967), Oller (1972), and McKenna (1976) found little difference between the two scoring methods in terms of validity and reliability. Bormuth (1969) suggested that exact scoring was, for most purposes, the most exact and economical. Russell (1978) argued that synonyms should not be accepted because: a) the research established did not support it, b) the criterion for cloze procedure scores was on only exact responses , c) scoring became subjective when synonyms were used.

This concern for the elimination of subjective judgement was also expressed by Culhane (1970), Walter

(1974), and Bortnick and Lopardo (1976). These findings are supportive of Taylor (1956) who stated "...the easiest ways of applying cloze procedure may be best for most uses", and also there is "...no advantage to putting oneself to the trouble of judging and scoring synonyms" (p.48).

However, Schoelles (1971) Bortnick and Lopardo (1976), and McKenna (1976) have demonstrated that for diagnostic purposes in individual student assessment or for the purposes of teaching(Rankin, 1964), the scoring of synonyms can prove useful. Both McKenna (1976) and Asher (1976) noted that high achieving children scored slightly higher when synonyms were counted.

In an unmodified cloze procedure, synonyms are not counted, but spelling errors, (when it is obvious the student has mispelled the correct word) are counted as correct. Improper word endings, however, should be counted as incorrect as Myers (1976) suggests that this indicates the student is not aware of the complete meaning of the sentence. Finally, the raw score for each student is the number of exact word replacements. The percentage score is calculated from the raw score and the total possible deletions.

The results of research related to scoring was summarized by Jongsma (1971) who stated:

The literature consistently shows the scoring

of exact replacements to be the most objective, efficient, and useful scoring system to use with the cloze procedure. Although slightly higher reliability has been obtained, at times, by using other procedures such as synonym count, the increased time and subjectivity necessary for such systems do not warrant their use. The exception to the synonym usage may be using the cloze procedure as a teaching technique. (p. 7-8)

Cloze Procedure Tests Validated Against Readability Formulas, Multiple-choice Tests and Standardized Tests

In order for the cloze procedure to gain recognition as a device for measuring reading comprehension it was necessary for this procedure, like readability formulas before it, to be validated against an established measure of reading comprehension. Bormuth (1967), cited Frederick (1955), Betts (1954), Flesch (1948), and Dale and Chall (1948) in demonstrating that the multiple-choice comprehension test was a "widely known frame of reference accepted in both readability research and in classroom practice" (p.292). Bormuth further explained that when a student correctly answered 75 - 90 percent of questions over a

passage the material was considered suitable for supervised instruction. Scores above 90 percent indicated materials might be used for independent study. Scores below 75 percent indicated the material was too difficult for normal instructional purposes (p.292).

First attempts to validate the cloze procedure were made by Taylor (1953) against the Flesch and Dale-Chall readability formulas which had themselves been validated against comprehension questions. Taylor found the cloze procedure consistently ranked selected passages in the same order as the two formulas and that the cloze procedure handled concept load more adequately. As Thelen (1974) pointed out "unlike readability formula, the cloze procedure evaluates the student's ability to handle the text" (p. 26). Subsequent research has most often used multiple-choice tests to validate cloze procedure scores. Bormuth (1968b) demonstrated the usefulness of this approach when he pointed out that "...studies seem to show that cloze and conventional tests measure the same process" (p.431). Taylor (1953) found a correlation of .76 between scores on a cloze procedure test and scores on a multiple-choice test made over the same passage. Bormuth (1968) cited Bormuth (1967) who found correlations of .73 to .84 between cloze and conventional tests (constructed by Bormuth) over the same passages. When corrections for

unreliabilities were made the correlations approached 1.00. Bormuth (1962) found a correlation of .92 between cloze procedure results and multiple-choice tests over the same passage. In a later study, Bormuth (1967), used four forms of the Gray Oral Reading Paragraphs and found correlations of .90 to .95 between cloze procedure scores and word recognition difficulties in the paragraphs and correlations of .91 to .96 between the cloze procedure and comprehension difficulties.

The cloze procedure was validated against the established multiple-choice comprehension test, but dissatisfaction has been expressed with the multiple-choice test itself and the cloze procedure has been looked to as a possible alternative to this measure of readability or comprehension. Bormuth (1963), (1968), Pennock (1973), Boyce (1974), Miller (1975), and Bortnick and Lopardo (1976) pointed out that with multiple-choice or other forms of comprehension questions, it was difficult to determine if the student's score reflected the difficulty of the passage, the difficulty of the questions, the student's difficulty in handling the questions or the subjectivity of the marker. Also it was difficult to know if the questions adequately sampled the content of the passage. The time required to construct and scrutinize a comprehension test in order to minimize the

aforementioned problems, is beyond the time constraints of most practitioners. This led Taylor (1957) to conclude that:

> Although cloze and comprehension tests were generally similar in the kinds of results they yielded, the two kinds of tests were very different in cost, effort, and time required for construction. The advantages seem to be with the cloze procedure in general, and the 'any' method of mutilation in particular. (p.25)

Bortnick and Lopardo (1976) pointed out that a cloze procedure test was objectively derived which allowed "different test writers to produce reliable and equivalent instruments over the same material" (p.116). The literature, then, appears to indicate that the cloze procedure is a more reliable, and therefore a superior, measure of comprehension (Bormuth 1963 , Miller and Coleman 1967 , Bormuth 1969b).

Having established the validity of the cloze procedure and its suggested superiority over comprehension questions based on a passage, researchers such as Bormuth and Coleman have now begun developing readability formulas validated against test results using the cloze procedure. It appears that more investigation is needed before it can be determined if

readability formulas derived from cloze procedure are more or less valid than multiple-choice derived formulas.

The correlation of cloze procedure results with results on standardized tests is of vital interest to this study which has used a standardized test as a measure of the students' reading grade level. Bormuth (1963) cited Fletcher (1955) and Rankin (1957) who found significant correlations between cloze procedure and an assortment of standardized reading tests. Jones and Pikulski (1974), found a correlation of .73 between the cloze procedure and the Comprehensive Test of Basic Skills. Smith and Zink (1977) found a correlation .74 between the total scores of the Davis Reading Test Form 2A and cloze procedure scores made over the same passages. They therefore reported that "The high correlation between scores on the DRT and the cloze test derived from the same instrument indicates that the cloze test measured the construct reading comprehension as measured by the standardized test" (p.397). Other studies by Jenkinson (1957), Ruddell (1963), Friedman (1964), (cited in Rankin 1965); Bormuth (1965), Tinzmann and Thompson (1977) found correlations that ranged from approximately .70 to .85. Weaver and Kingston (1963), using the total raw score of the Davis Reading Test, found a low correlation between cloze

procedure results and standardized tests in respect to "
verbal comprehension." They found that the abilities
required to complete a cloze procedure were related to
redundancy utilization. Rankin (1965) pointed out that
only the Weaver and Kingston study found a low
correlation between the cloze and standardized test.
Bormuth (1969) cautioned that the data used by Weaver
and Kingston should be questioned on several accounts.

Research seems to indicate that there is a
significant correlation between cloze procedure scores
and scores on standardized tests. This being the case,
many would argue that the cloze is the preferable
testing device. Jones and Pikulski (1974) pointed out
that the accuracy of standardized tests in precisely
identifying reading achievement was questionable.
Bortnick and Lopardo (1976) explained that "Standardized
tests are limited to normative interpretations, which
permits only comparison of one group or individual with
the norm population" (p.114). Rakes and McWilliams
(1978) also pointed out that cloze procedure or other
informal tests are less expensive than standardized test
batteries. Taking into consideration the aforementioned
criticisms it appears that the literature is less
critical of standardized tests than of non-standardized
multiple-choice or completion comprehension questions.
It would appear advisable to use the former when

attempting to establish the validity of the cloze procedure over a given passage.

Frame of Reference for Cloze Procedure Scores

For some time the main weakness of the cloze procedure as a measure of readability was the absence of criteria for interpreting raw scores. The relative difficulty of two or more passages could be determined but no interpretation could be placed upon the difficulty of each passage. (Rankin 1970 cited in Van Rooy 1973 p. 7)

In an attempt to establish such criteria, standards set by Thorndike (1917), and Betts (1954) have been accepted. They indicated that materials were at a child's instructional level when he/she was able to answer correctly 75 percent of the questions asked him/her about the passage, and at his/her independent level when he/she could answer 90 percent. Materials on which the child scored below 75 percent were considered too difficult for instructional purposes. In his first attempt to establish a frame of reference for cloze scores Bormuth (1967) compared cloze procedure and multiple-choice comprehension tests administered over

the same passages and to the same readers. He found a cloze score of 38 percent was comparable to a multiple-choice score of 75 percent and a cloze score of 50 percent was comparable to a multiple-choice score of 90 percent. He cautioned that when multiple-choice scores were corrected for guessing, a cloze score of 43 was required to reach the 75 percent multiple-choice criterion. In this 1967 study Bormuth observed ceiling effects on the multiple-choice scores which may have led to the low cloze scores when compared to the multiple-choice scores. The following year Bormuth undertook a further investigation, this time using the Gray-Oral Reading Tests . "Two of the four paragraphs on each level of difficulty were randomly assigned to each subject who took these two paragraphs as cloze tests. The complementary pair of paragraphs was taken by the same subject as oral reading tests" (Bormuth 1968). In this study Bormuth found cloze procedure scores of 44 percent and 57 percent comparable to comprehension criterion scores of 75 and 90 percent. Bormuth appears to view the 1968 study as the most reliable for he quoted these results in subsequent papers. Rankin and Culhane (1969) replicated the Bormuth study and found cloze procedure scores of 41 percent and 61 percent respectively, comparable to criterion scores of 75 and 90 percent. This led Rankin and Culhane to conclude

that the 1968 Bormuth scores were valid. Because these results vary slightly the literature seems to recommend that students whose scores fall between 40 and 59 percent would profit from instruction on that material whereas students scoring below 40 percent would find the material too difficult for instruction. Those scoring sixty or above would find the material suitable for independent study . Pennock (1973) and Dishner (1973) reported that students scoring above 65 percent were likely to gain little new information from that material.

With the criteria for interpreting raw cloze scores now in place the practitioner can have some degree of confidence in determining the suitability of materials for a given student.

Criticisms of the Cloze Procedure

Although the literature appears to support both the validity and applicability of cloze procedure, it has not gone without criticism . The major criticism, or caution to be considered, concerns the fact that any cloze procedure test constructed over a given passage cannot be assumed to be of the same difficulty as any other cloze procedure test constructed over the same passage. If an every fifth word deletion system is used

there are five possible cloze procedure tests. If an
every tenth word deletion system is used there are ten
possible tests and so on. This concern was expressed by
both Bormuth (1964) and Boyce (1974). Bormuth found the
longer the test the less variability occurred but he
suggested that for research purposes, more than one test
form be used. Boyce explained that the variability was
not a problem if the test was being used to rank
students but it might pose problems when the score was
used to compare a student's score to an established
criterion score and might result in an incorrect
decision as to the suitability of material for a
student. Boyce (1978) found that the length of a word
had a definite influence on the student's ability to
replace the word. The mean replacement rate for one and
two syllable words was 73.4 percent whereas the
replacement rate for words seven letters or longer was
21.2 percent. Recognizing that word length is certainly
not the only factor affecting replacement ease, the
practitioner may still be well advised to use
professional judgement when selecting the passage to
make sure it is not weighted towards either long or
short words. It should be kept in mind that scores are
to be interpreted within very wide ranges labelled
frustration, instructional or independent. It does not
appear that they were meant to be rigidly compared to

criterion scores.

Other criticisms of the cloze procedure included studies by Sauer (1969, reported by Riley 1973) and Kirby (1967 cited by Walter 1974) who found that the cloze procedure did not adequately assess the reading levels of students in the lower elementary grades. Kirby (1968) found that students whose word recognition abilities were adequate performed significantly better on the cloze procedure than those with less developed recognition skills. Pollock (1974) compared cloze procedure scores to the informal reading inventory scores of subjects from both a middle and upper socioeconomic level and a lower socioeconomic level and concluded that the cloze procedure yielded depressed scores for students from lower socioeconomic levels and was therefore inappropriate for use with such students.

Finally Tuinman (1975) suggested that the cloze procedure measures local redundancy more than the comprehension of major ideas. These limitation in the cloze procedure are certainly useful background knowledge for the practitioner but they would not appear to be significant in stature nor significantly supported in the literature to in any way affect the confidence with which the cloze procedure might be used in an average classroom.

Modifications on the Cloze Procedure

Over the years various researchers have proposed numerous alterations or modifications to the cloze procedure. A streamlined sampling process, most useful for the practitioner, was proposed by Pennock (1973). He suggested that rather than prepare six to twelve passages to be tested on a sample population, that a readability formula be applied to the passages and the passage that came closest to the mean readability of all the passages should be prepared as a cloze procedure. Such a process would save hours of work and as such would increase the likelihood of the cloze procedure being used.

Most other proposals have varied more widely afield from the traditional cloze procedure. Hafner (1965) conducted a study using deletions of letters from words and found a high correlation between this test and reading results. Carver (1974) constructed a test in which every second word contained only the first letter. One in every five of these initial letters was replaced with an incorrect letter. Subjects were asked to make the corrections and fill in the blanks. Carver theorized that this type of test gave the reader a chunk of information to assist in the retrieval of the correct word. Although Carver indicated that further research was required he reported that the results suggested this

type of test was as valid as the cloze procedure and more reliable in measuring reading gain.

Cunningham and Cunningham (1978) compared the cloze procedure with a limited cloze procedure in which the deleted words were randomly ordered and placed above the passage. In study one they found the percentage range of 73 - 93 percent was equivalent to the instructional level range on the cloze procedure and in study two they found a range of 60 - 81 percent. They concluded that the limited cloze procedure was "as valid, reliable, objective and practical as regular cloze, but less interpretable" (p.211). Entin and Klare (1978) studied the implication of using a dash for each letter of the deleted word. Two sets of deletions were used in the study - the same two for the solid line and dash forms. This was done to minimize the possibility of hitting a single unrepresentative easy or difficult passage. Subjects were also given a multiple-choice test. As expected cloze procedure scores on the dash form were higher but their correlation with the multiple-choice scores was about the same as the standard format. Entin and Klare concluded that "the dash format should be at least as good a measure of comprehension as the standard format" (p.427).

Anderson (1971) and Spooncer (1974) found that when the standard length blank was replaced by a blank the

same size as the deleted word, the cloze procedure scores were not significantly increased. Boyce (1974) cited Anderson (1971) who further suggested cloze procedure passages could be constructed by gluing paper over the words in the original passage that were to be deleted. The passage could then be photocopied. Boyce suggested that this would give the student all the contextual clues available in regular reading. He further simplified the Anderson process by using liquid paper to delete the words. Because he found that the space left was often too small to allow the student to print the word, he numbered the blanks and provided a separate numbered answer sheet. Unfortunately, the validy of this method, as opposed to the cloze procedure, was not tested.

The most recent innovation in the cloze procedure was outlined by McCabe (1979) . The McCabe proposal is outlined in detail in Chapter I. Studies by Hafner (1965) who found that the deletion of letters had a high correlation with reading results, and Carver (1974) who supported the concept of giving the reader a chunk of information to assist in retrieval, lend credence to McCabe's proposal which involves the deletion of letters, partial words and whole words. It was the purpose of this research to determine the validy of this approach in relation to the <u>Stanford Diagnostic.</u>

CHAPTER III

A DESCRIPTION OF THE STUDY

The entire population of grade nine students in a suburban district in British Columbia was tested for reading achievement as measured by the Stanford Diagnostic Reading Test, Blue Level, Form A (hereafter referred to as Stanford Diagnostic). A passage was then selected using the Fry Readability Graph. A cloze procedure, a Beginning of the Page Procedure (B.O.P.P.) and an "instant" Beginning of the Page Procedure were constructed over this same passage. The cloze procedure and modified cloze procedure tests were then distributed randomly to every third student. Students were instructed to fill in the blanks with the exact word, words or parts of words which had been deleted. Only exact replacements of deletions were scored.

Selection of Subjects

Subjects tested were grade nine students in a middle class suburban district. The area was a working class district with a high percentage of single parent families. The percentage of immigrant families was

minute. Testing took place in April within the English classroom as all grade nine students took English over the entire year. Of the two hundred and thirty-nine students who took part in testing only 196 scores could be considered due to absenteeism on either of the testing days. One hundred and six of these subjects were male and 90 were female.

Procedures for Administering and Scoring the Stanford Diagnostic Reading Test

The Stanford Diagnostic was designed to measure the reading capabilities of students in grades 9 through 12 plus college. It was designed to provide particularly accurate assessment of low-achieving students but did not appear to give an equally accurate assessment of superior readers. Over a period of one week, all grade 9 students were administered the comprehension section of the Stanford Diagnostic. Each student was given a test booklet and an answer sheet. They were instructed to darken in the circle corresponding to the answer they chose. Thirty-five minutes was allowed for the administration of the comprehension subtest and this was strictly adhered to. When time had expired, tests were collected and hand scored using an answer stencil. The total comprehension raw score was computed.

Procedures for Selecting the Passage

The passage was selected from the Barnell Loft, Specific Skills Series, Book I, "Getting the Facts". This book is recommended for students working at the grade 9 instructional level. Passages in the book were assessed in order to find a selection reasonably free of proper nouns and numbers. The Fry Readability Graph was then applied to two one hundred word segments of the passage and both were found to have a readability level of grade 7.5. As the Fry Graph measures frustration to independent level it was felt that this passage was suitable for the instruction of an average student in the eighth month of grade 9.

Procedures for Constructing, Administering and Marking the Cloze Procedure

A fifty item cloze procedure test was constructed using a selection found suitable by the Fry Graph. The first and last sentences were left intact and as suggested by Taylor (1956), Bormuth (1968) and others, an every fifth word deletion pattern was used. The standard length space was employed as this appeared to be the most commonly used procedure (Taylor 1956, 1957, Bormuth 1967, 1968, 1969, Bortnick and Lopardo 1976) although Anderson (1971) and Spooncer (1974) found the

length of the space made no significant difference to test results.

Students were given the cloze test in their English classes within one week of taking the Stanford Diagnostic. The cloze procedure was given randomly to every third student who was instructed to put his/her name on his/her paper. This was to allow correlation of cloze procedure results to standardized test results and also to create a seriousness often absent when names are not required. Using what Rankin (1965) called the pre-cloze technique, students were asked to fill in the deleted words without having read the unmutilated passage. It was explained that they were not expected to be able to fill in all the spaces but that a score of just twenty out of fifty was equivalent to seventy-five percent on a multiple-choice exam. Students were told that they could take as much time as they required to complete the test.

The cloze procedure tests were hand scored and only exact replacements of deleted words were accepted. Minor spelling errors, where it was clear that the deleted word was intended , were accepted. Scores were multiplied by two to obtain a percentage. (see appendix G for the cloze procedure)

Procedures for Constructing, Administering and Marking the Beginning of the Page Procedure

The passage used for the standard cloze procedure test was typed onto a 9 1/2 by 11 inch (24 cm high by 27.5 cm wide) piece of white paper. The typewriter was set for one and a half spaces between lines. A strip of paper 1/10th the length of the average line was then glued one inch (2.5 cm) from the left hand margin. The first and last sentences were left intact so the strip of paper did not intersect the lines containing these sentences. As noted by Boyce (1974), the space left when a typewritten word is deleted is often too small to allow a student to reproduce the word. Each deletion was therefore numbered and a space with the corresponding number was provided at the end of each line. This master B.O.P.P. was then photocopied to provide the required number of tests.

The B.O.P.P. was randomly distributed to another one-third of the students who were also required to write their names on the papers. They were instructed that a word, words or parts of words were missing and on the space provided in the right hand margin, they were to write in the exact words that had been deleted. Again the students were given as much time as they required and again the pre-cloze procedure was employed.

The scoring procedure was not discussed by McCabe

but it was decided that only exact replacements and
reasonable spelling errors would be credited. One mark
was given for each partial word replaced and two marks
were given for every whole word replaced. The students'
points were added as were the total possible test points
and a percentage score was calculated for each student.
(see appendix H for the B.O.P.P.)

Procedures for Constructing, Administering and Marking
the " Instant " Beginning of the Page Procedure

The passage used for the cloze procedure and the
B.O.P.P. was also used for the "instant" B.O.P.P. The
selection was photocopied, and a strip of paper one-
tenth the length of the average line was glued 1 inch
(2.5 cm) from the left margin, leaving the first and
last sentences intact. The deletions were once again
numbered and a uniform blank space with the
corresponding number was provided in the right hand
margin. (appendix I) The resulting test was photocopied
to provide the required number of tests for the final
1/3 of the experimental population.

The "instant" B.O.P.P. like the cloze procedure and
the B.O.P.P., was administered during the English period
and within one week of taking the Stanford Diagnostic.
Students were given as much time as they required and

the pre-cloze procedure was employed. The instructions to students were the same as those given for the B.O.P.P. and the scoring procedures were also the same. In both the B.O.P.P. and the "instant" B.O.P.P. as in the cloze procedure, students were encouraged to record their reactions to the test.

Analysis of the Data

1. The mean and standard deviation was calculated for the raw scores on the Stanford Diagnostic for each of the three test groups, cloze procedure, B.O.P.P. and "instant" B.O.P.P. Means and standard deviations were also calculated for the total male and female populations. (Tables I - II)

2. A table is provided demonstrating the distribution of subjects by sex for each cell. (Table IV)

3. A one way analysis of variance (an F Statistic) was calculated to determine if there was a statistically significant sex difference for Stanford Diagnostic mean scores for the total population and for groups cloze procedure, B.O.P.P. and "instant" B.O.P.P. (Table III)

4. Histograms were prepared for the total population and for subgroups cloze procedure, B.O.P.P. and "instant" B.O.P.P., based on raw scores from the Stanford Diagnostic. The percent scores for groups cloze procedure, B.O.P.P. and "instant" B.O.P.P. were also presented in histograms. (Figures I - VII)

5. The mean and standard deviation was prepared for the percentage scores on each of the subgroups, cloze procedure, B.O.P.P. and "instant" B.O.P.P. The mean for sex and the significance of the difference between means was also calculated. (Tables V, VI, VII)

6. An estimated equivalency graph was prepared, comparing scores on the cloze procedure, the B.O.P.P. and the "instant" B.O.P.P. to both the raw scores and the grade scores on the Stanford Diagnostic. (Table IX)

7. A prediction equation and a corresponding scatter plot was prepared for each of the groups cloze procedure, B.O.P.P. and "instant" B.O.P.P. using each group as the criterion and the Stanford Diagnostic Test as the predictor. (Figure VIII - X)

8. The predicted regression lines for each group cloze procedure, B.O.P.P. and "instant" B.O.P.P. were drawn on a single graph. (Figure XI)

9. Pearson, product-moment correlations were computed for cloze procedure, B.O.P.P. and "instant" B.O.P.P. percentage scores with Stanford Diagnostic raw scores. (Table X)

10. The significance of the correlations of the cloze procedure, B.O.P.P. and "instant" B.O.P.P. with the Stanford Diagnostic were computed. (Table XI)

FIGURE 1 STANFORD DIAGNOSTIC RAW SCORES FOR THE TOTAL POPULATION.

MIDPOINT	HIST%	COUNT FOR 1=STANFORD (EACH X=1)	
9.000	.5	1	+X
15.000	.5	1	+X
16.000	0.	0	+
17.000	0.	0	+
18.000	1.0	2	+XX
19.000	.5	1	+X
20.000	0.	0	+
21.000	1.0	2	+XX
22.000	1.0	2	+XX
23.000	1.5	3	+XXX
24.000	1.0	2	+XX
25.000	1.0	2	+XX
26.000	3.1	6	+XXXXXX
27.000	3.6	7	+XXXXXXX
28.000	2.6	5	+XXXXX
29.000	2.0	4	+XXXX
30.000	2.0	4	+XXXX
31.000	5.6	11	+XXXXXXXXXXX
32.000	2.0	4	+XXXX
33.000	2.6	5	+XXXXX
34.000	2.6	5	+XXXXX
35.000	2.0	4	+XXXX
36.000	4.6	9	+XXXXXXXXX
37.000	5.1	10	+XXXXXXXXXX
38.000	1.5	3	+XXX
39.000	4.6	9	+XXXXXXXXX
40.000	3.6	7	+XXXXXXX
41.000	4.1	8	+XXXXXXXX
42.000	2.0	4	+XXXX
43.000	1.0	2	+XX
44.000	3.1	6	+XXXXXX
45.000	3.6	7	+XXXXXXX
46.000	2.6	5	+XXXXX
47.000	3.1	6	+XXXXXX
48.000	2.6	5	+XXXXX
49.000	1.0	2	+XX
50.000	4.6	9	+XXXXXXXXX
51.000	2.6	5	+XXXXX
52.000	3.6	7	+XXXXXXX
53.000	3.6	7	+XXXXXXX
54.000	3.6	7	+XXXXXXX
55.000	2.0	4	+XXXX
56.000	1.0	2	+XX
57.000	0.	0	+
58.000	.5	1	+X
TOTAL		196	(INTERVAL WIDTH = 1.0000)

FIGURE 1

HISTOGRAM

MIDPOINT   HIST%   COUNT FOR 1.STANFORD   (EACH X= 1)

```
   0.          0.        0 +
   6.000       0.        0 +
  12.000        .5       1 +X
  18.000       2.0       4 +XXXX
  24.000       8.7      17 +XXXXXXXXXXXXXXXXX
  30.000      17.9      35 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  36.000      18.4      36 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  42.000      18.4      36 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  48.000      17.3      34 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  54.000      16.3      32 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
  60.000        .5       1 +X
```

TOTAL   196   (INTERVAL WIDTH = 6.0000)
FIGURE 1   STANFORD DIAGNOSTIC RAW SCORES FOR THE
           TOTAL POPULATION

FIGURE 2   STANFORD DIAGNOSTIC RAW SCORES FOR THE
             SUBGROUP CLOZE PROCEDURE

HISTOGRAM   <1> TREATMENT:CLOZE PROCEDURE

MIDPOINT   HIST%   COUNT FOR 1.STANFORD   (EACH X= 1)

```
15.000       1.6       1 +X
16.000       0.        0 +
17.000       0.        0 +
18.000       1.6       1 +X
19.000       0.        0 +
20.000       0.        0 +
21.000       0.        0 +
22.000       1.6       1 +X
23.000       1.6       1 +X
24.000       0.        0 +
25.000       1.6       1 +X
26.000       1.6       1 +X
27.000       4.7       3 +XXX
28.000       3.1       2 +XX
29.000       1.6       1 +X
30.000       0.        0 +
31.000       9.4       6 +XXXXXX
32.000       4.7       3 +XXX
33.000       4.7       3 +XXX
34.000       3.1       2 +XX
35.000       4.7       3 +XXX
36.000       1.6       1 +X
37.000       4.7       3 +XXX
38.000       1.6       1 +X
39.000       7.8       5 +XXXXX
40.000       1.6       1 +X
41.000       1.6       1 +X
42.000       1.6       1 +X
43.000       1.6       1 +X
44.000       1.6       1 +X
45.000       0.        0 +
46.000       4.7       3 +XXX
47.000       1.6       1 +X
48.000       3.1       2 +XX
49.000       1.6       1 +X
50.000       6.3       4 +XXXX
51.000       0.        0 +
52.000       4.7       3 +XXX
53.000       4.7       3 +XXX
54.000       1.6       1 +X
55.000       3.1       2 +XX
```

TOTAL   64 (INTERVAL WIDTH = 1.0000)

FIGURE 2

HISTOGRAM   <1> TREATMENT:CLOZE PROCEDURE

MIDPOINT   HIST%   COUNT FOR 1.STANFORD   (EACH X= 1)

```
    0.          0.        0 +
    6.000       0.        0 +
   12.000       0.        0 +
   18.000       3.1       2 +XX
   24.000       6.3       4 +XXXX
   30.000      23.4      15 +XXXXXXXXXXXXXXX
   36.000      20.3      13 +XXXXXXXXXXXXX
   42.000      15.6      10 +XXXXXXXXXX
   48.000      17.2      11 +XXXXXXXXXXX
   54.000      14.1       9 +XXXXXXXXX
   60.000       0.        0 +
```

TOTAL   64   (INTERVAL WIDTH = 6.0000)
FIGURE 2 STANFORD DIAGNOSTIC RAW SCORES FOR THE
         SUBGROUP CLOZE PROCEDURE

FIGURE 3 <u>STANFORD DIAGNOSTIC</u> RAW SCORES FOR THE
        SUBGROUP B.O.P.P.

HISTOGRAM   <2> TREATMENT:B.O.P.P.

MIDPOINT   HIST%   COUNT FOR 1.STANFORD   (EACH X= 1)

MIDPOINT	HIST%	COUNT	
19.000	1.5	1	+X
20.000	0.	0	+
21.000	1.5	1	+X
22.000	1.5	1	+X
23.000	0.	0	+
24.000	1.5	1	+X
25.000	1.5	1	+X
26.000	6.0	4	+XXXX
27.000	1.5	1	+X
28.000	4.5	3	+XXX
29.000	4.5	3	+XXX
30.000	4.5	3	+XXX
31.000	4.5	3	+XXX
32.000	1.5	1	+X
33.000	0.	0	+
34.000	3.0	2	+XX
35.000	0.	0	+
36.000	7.5	5	+XXXXX
37.000	3.0	2	+XX
38.000	1.5	1	+X
39.000	3.0	2	+XX
40.000	1.5	1	+X
41.000	3.0	2	+XX
42.000	1.5	1	+X
43.000	0.	0	+
44.000	1.5	1	+X
45.000	6.0	4	+XXXX
46.000	1.5	1	+X
47.000	3.0	2	+XX
48.000	3.0	2	+XX
49.000	1.5	1	+X
50.000	4.5	3	+XXX
51.000	6.0	4	+XXXX
52.000	6.0	4	+XXXX
53.000	3.0	2	+XX
54.000	3.0	2	+XX
55.000	1.5	1	+X
56.000	1.5	1	+X

TOTAL   67   (INTERVAL WIDTH = 1.0000)

FIGURE 3

```
HISTOGRAM    <2>  TREATMENT:B.O.P.P.

MIDPOINT     HIST%   COUNT FOR 1.STANFORD   (EACH X= 1)

   0.          0.      0 +
   6.000       0.      0 +
  12.000       0.      0 +
  18.000       1.5     1 +X
  24.000      11.9     8 +XXXXXXXX
  30.000      20.9    14 +XXXXXXXXXXXXXX
  36.000      14.9    10 +XXXXXXXXXX
  42.000      10.4     7 +XXXXXXX
  48.000      19.4    13 +XXXXXXXXXXXXX
  54.000      20.9    14 +XXXXXXXXXXXXXX
  60.000       0.      0 +

TOTAL                67  (INTERVAL WIDTH = 6.0000)
```
FIGURE 3 STANFORD DIAGNOSTIC RAW SCORES FOR THE
         SUBGROUP B.O.P.P.

FIGURE 4 <u>STANFORD DIAGNOSTIC</u> RAW SCORES FOR THE
          SUBGROUP "INSTANT" B.O.P.P.

HISTOGRAM   <3> TREATMENT:"INSTANT" B.C.P.P.

MIDPOINT   HIST%   COUNT FOR 1.STANFORD   (EACH X= 1)

```
 9.000      1.5     1 +X
18.000      1.5     1 +X
19.000      0.      0 +
20.000      0.      0 +
21.000      1.5     1 +X
22.000      0.      0 +
23.000      3.1     2 +XX
24.000      1.5     1 +X
25.000      0.      0 +
26.000      1.5     1 +X
27.000      4.6     3 +XXX
28.000      0.      0 +
29.000      0.      0 +
30.000      1.5     1 +X
31.000      3.1     2 +XX
32.000      0.      0 +
33.000      3.1     2 +XX
34.000      1.5     1 +X
35.000      1.5     1 +X
36.000      4.6     3 +XXX
37.000      7.7     5 +XXXXX
38.000      1.5     1 +X
39.000      3.1     2 +XX
40.000      7.7     5 +XXXXX
41.000      7.7     5 +XXXXX
42.000      3.1     2 +XX
43.000      1.5     1 +X
44.000      6.2     4 +XXXX
45.000      4.6     3 +XXX
46.000      1.5     1 +X
47.000      4.6     3 +XXX
48.000      1.5     1 +X
49.000      0.      0 +
50.000      3.1     2 +XX
51.000      1.5     1 +X
52.000      0.      0 +
53.000      3.1     2 +XX
54.000      6.2     4 +XXXX
55.000      1.5     1 +X
56.000      1.5     1 +X
57.000      0.      0 +
58.000      1.5     1 +X
```

TOTAL   65 (INTERVAL WIDTH = 1.0000)

FIGURE 4

HISTOGRAM   <3> TREATMENT:"INSTANT" B.O.P.P.

MIDPOINT   HIST%   COUNT FOR 1.STANFORD   (EACH X= 1)

```
     0.           0.         0 +
     6.000        0.         0 +
    12.000        1.5        1 +X
    18.000        1.5        1 +X
    24.000        7.7        5 +XXXXX
    30.000        9.2        6 +XXXXXX
    36.000       20.0       13 +XXXXXXXXXXXXX
    42.000       29.2       19 +XXXXXXXXXXXXXXXXXXX
    48.000       15.4       10 +XXXXXXXXXX
    54.000       13.8        9 +XXXXXXXXX
    60.000        1.5        1 +X
```

TOTAL   65   (INTERVAL WIDTH = 6.0000)
FIGURE 4 STANFORD DIAGNOSTIC RAW SCORES FOR THE
          SUBGROUP "INSTANT" B.O.P.P.

FIGURE 5 PERCENT SCORES FOR THE GROUP CLOZE
PROCEDURE

MIDPOINT	HIST%	COUNT	FOR 5.PERCENT	(EACH X= 1)
10.000	1.6	1	+X	
12.000	0.	0	+	
14.000	1.6	1	+X	
16.000	0.	0	+	
18.000	3.1	2	+XX	
20.000	3.1	2	+XX	
22.000	3.1	2	+XX	
24.000	3.1	2	+XX	
26.000	3.1	2	+XX	
28.000	1.6	1	+X	
30.000	1.6	1	+X	
32.000	9.4	6	+XXXXXX	
34.000	1.6	1	+X	
36.000	4.7	3	+XXX	
38.000	14.1	9	+XXXXXXXXX	
40.000	6.3	4	+XXXX	
42.000	3.1	2	+XX	
44.000	4.7	3	+XXX	
46.000	9.4	6	+XXXXXX	
48.000	4.7	3	+XXX	
50.000	3.1	2	+XX	
52.000	0.	0	+	
54.000	6.3	4	+XXXX	
56.000	3.1	2	+XX	
58.000	3.1	2	+XX	
60.000	0.	0	+	
62.000	0.	0	+	
64.000	0.	0	+	
66.000	3.1	2	+XX	
68.000	0.	0	+	
70.000	0.	0	+	
72.000	1.6	1	+X	

TOTAL   64   (INTERVAL WIDTH = 2.0000)

FIGURE 5

```
HISTOGRAM   <1> TREATMENT:CLOZE PROCEDURE
MIDPOINT    HIST%   COUNT FOR 5.PERCENT   (EACH X= 1)
   0.         0.       0 +
 10.000       3.1      2 +XX
 20.000      12.5      8 +XXXXXXX
 30.000      17.2     11 +XXXXXXXXXX
 40.000      32.8     21 +XXXXXXXXXXXXXXXXXXXX
 50.000      23.4     15 +XXXXXXXXXXXXXX
 60.000       6.3      4 +XXX
 70.000       4.7      3 +XX
 80.000       0.       0 +
 90.000       0.       0 +
100.00        0.       0 +

TOTAL                 64  (INTERVAL WIDTH = 10.000)
FIGURE 5    PERCENT SCORES FOR THE GROUP CLOZE
            PROCEDURE
```

FIGURE 6   PERCENT SCORES FOR THE GROUP B.O.P.P.

MIDPOINT   HIST%   CCUNT FOR 5.PERCENT   (EACH X= 1)

MIDPOINT	HIST%	CCUNT	
14.000	1.5	1	+X
16.000	0.	0	+
18.000	3.0	2	+XX
20.000	1.5	1	+X
22.000	1.5	1	+X
24.000	1.5	1	+X
26.000	0.	0	+
28.000	3.0	2	+XX
30.000	3.0	2	+XX
32.000	3.0	2	+XX
34.000	3.0	2	+XX
36.000	0.	0	+
38.000	7.5	5	+XXXXX
40.000	3.0	2	+XX
42.000	3.0	2	+XX
44.000	0.	0	+
46.000	4.5	3	+XXX
48.000	4.5	3	+XXX
50.000	3.0	2	+XX
52.000	1.5	1	+X
54.000	4.5	3	+XXX
56.000	1.5	1	+X
58.000	3.0	2	+XX
60.000	1.5	1	+X
62.000	4.5	3	+XXX
64.000	0.	0	+
66.000	7.5	5	+XXXXX
68.000	3.0	2	+XX
70.000	3.0	2	+XX
72.000	1.5	1	+X
74.000	1.5	1	+X
76.000	6.0	4	+XXXX
78.000	3.0	2	+XX
80.000	3.0	2	+XX
82.000	1.5	1	+X
84.000	3.0	2	+XX
86.000	0.	0	+
88.000	0.	0	+
90.000	3.0	2	+XX
92.000	0.	0	+
94.000	1.5	1	+X

TOTAL   67   (INTERVAL WIDTH = 2.0000)

FIGURE 6

HISTOGRAM   <2> TREATMENT:B.O.P.P.

MIDPOINT   HIST%   COUNT FOR 5.PERCENT   (EACH X= 1)

0.	0.	0	+
10.000	1.5	1	+ X.
20.000	7.5	5	+ XXXXX
30.000	11.9	8	+ XXXXXXXX
40.000	13.4	9	+ XXXXXXXXX
50.000	17.9	12	+ XXXXXXXXXXXX
60.000	10.4	7	+ XXXXXXX
70.000	16.4	11	+ XXXXXXXXXXX
80.000	16.4	11	+ XXXXXXXXXXX
90.000	4.5	3	+ XXX
100.00	0.	0	+

TOTAL   67   (INTERVAL WIDTH = 10.000)
FIGURE 6   PERCENT SCORES FOR THE GROUP B.O.P.P.

FIGURE 7 PERCENT SCORES FOR THE GROUP
          "INSTANT" B.O.P.P.

MIDPOINT   HIST%   COUNT FOR 5.PERCENT   (EACH X= 1)

```
10.000      3.1     2 +XX
12.000      6.2     4 +XXXX
14.000      0.      0 +
16.000      0.      0 +
18.000      3.1     2 +XX
20.000      0.      0 +
22.000      1.5     1 +X
24.000      0.      0 +
26.000      0.      0 +
28.000      4.6     3 +XXX
30.000      0.      0 +
32.000      4.6     3 +XXX
34.000      4.6     3 +XXX
36.000      1.5     1 +X
38.000      1.5     1 +X
40.000      1.5     1 +X
42.000      0.      0 +
44.000      4.6     3 +XXX
46.000      0.      0 +
48.000      0.      0 +
50.000      4.6     3 +XXX
52.000      6.2     4 +XXXX
54.000      4.6     3 +XXX
56.000      6.2     4 +XXXX
58.000      6.2     4 +XXXX
60.000      3.1     2 +XX
62.000      0.      0 +
64.000      3.1     2 +XX
66.000      9.2     6 +XXXXXX
68.000      3.1     2 +XX
70.000      4.6     3 +XXX
72.000      1.5     1 +X
74.000      1.5     1 +X
76.000      1.5     1 +X
78.000      3.1     2 +XX
80.000      1.5     1 +X
82.000      0.      0 +
84.000      1.5     1 +X
86.000      0.      0 +
88.000      1.5     1 +X
```

TOTAL   65   (INTERVAL WIDTH = 2.0000)

FIGURE 7

```
HISTOGRAM    <3>  TREATMENT:"INSTANT" B.O.P.P.

MIDPOINT     HIST%   COUNT FOR 5.PERCENT   (EACH X= 1)

    0.         0.        0 +
   10.000      9.2       6 +XXXXXX
   20.000      4.6       3 +XXX
   30.000     13.8       9 +XXXXXXXXX
   40.000      9.2       6 +XXXXXX
   50.000     15.4      10 +XXXXXXXXXX
   60.000     18.5      12 +XXXXXXXXXXXX
   70.000     20.0      13 +XXXXXXXXXXXXX
   80.000      7.7       5 +XXXXX
   90.000      1.5       1 +X
  100.00       0.        0 +

  TOTAL                 65  (INTERVAL WIDTH = 10.000)
  FIGURE 7  PERCENT SCORES FOR THE GROUP
            "INSTANT" B.O.P.P.
```

FIGURE 8 SCATTERGRAM OF CLOZE PROCEDURE PERCENT SCORES

AND STANFORD DIAGNOSTIC RAW SCORES
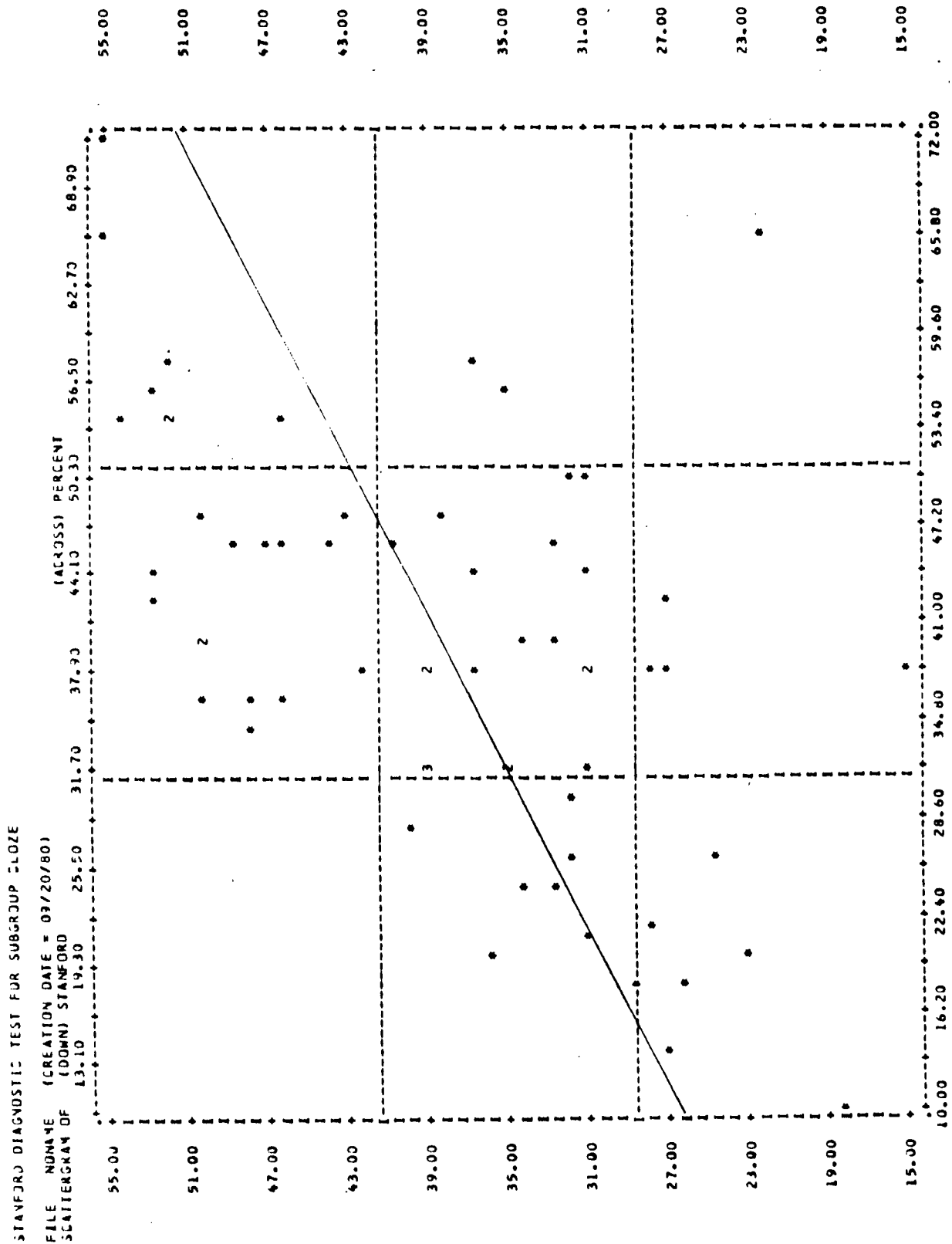
FIGURE 9 SCATTERGRAM OF B.O.P.P. PERCENT SCORES AND STANFORD DIAGNOSTIC RAW SCORES
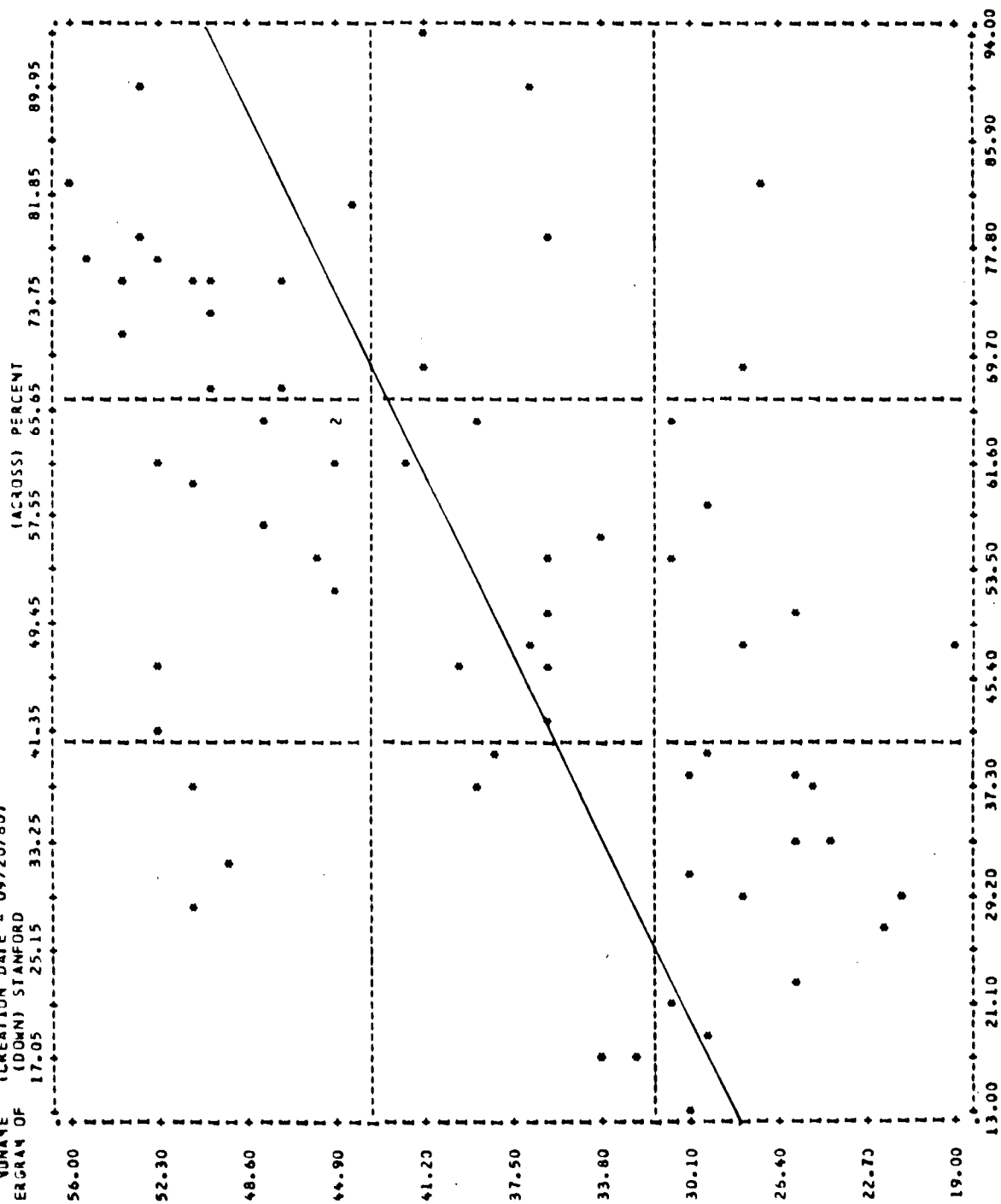
FIGURE 10 SCATTERGRAM OF "INSTANT" B.O.P.P. PERCENT

SCORES AND STANFORD DIAGNOSTIC RAW SCORES
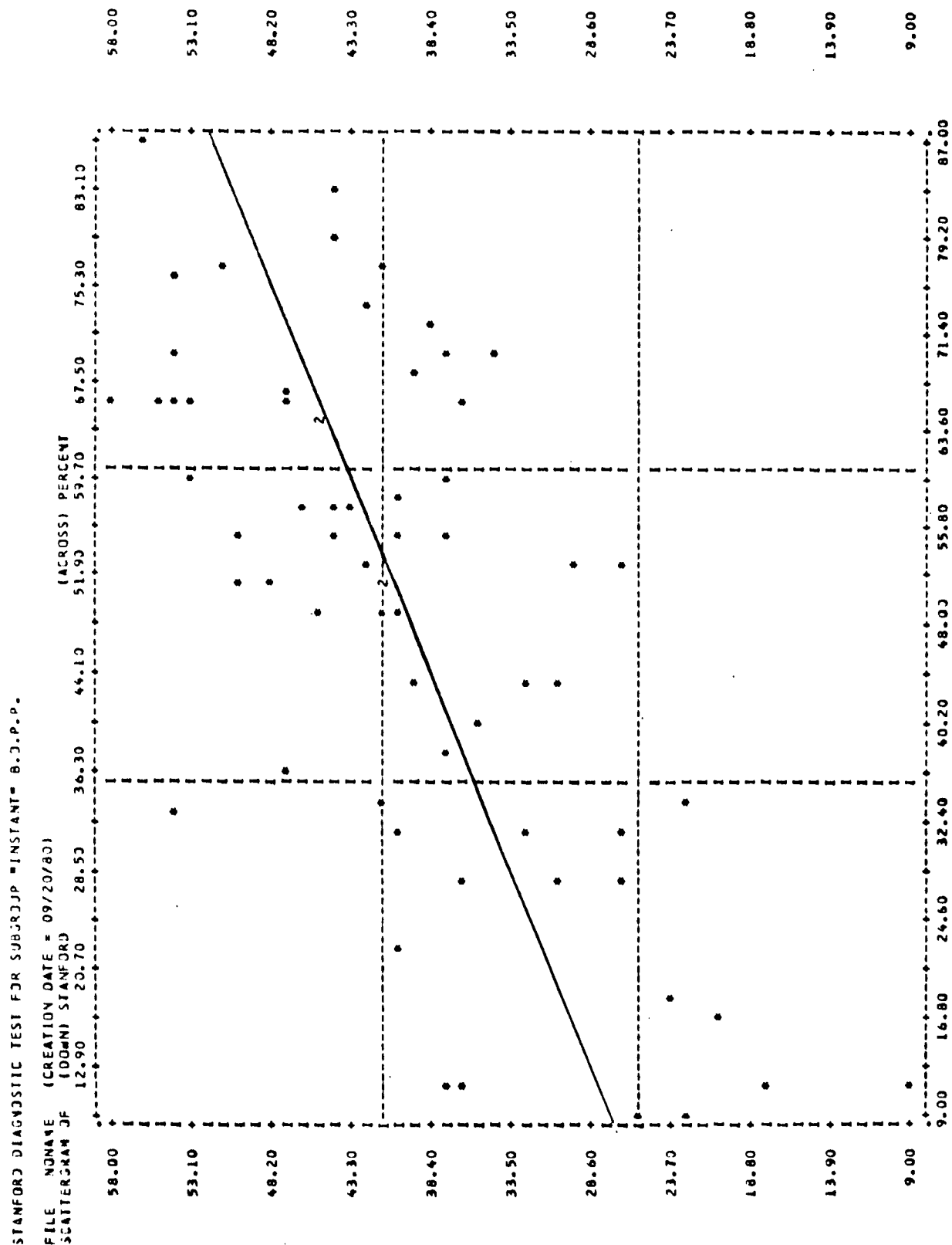
FIGURE 11 PREDICTED REGRESSION LINES FOR GROUPS CLOZE PROCEDURE, B.O.P.P. AND "INSTANT" B.O.P.P.
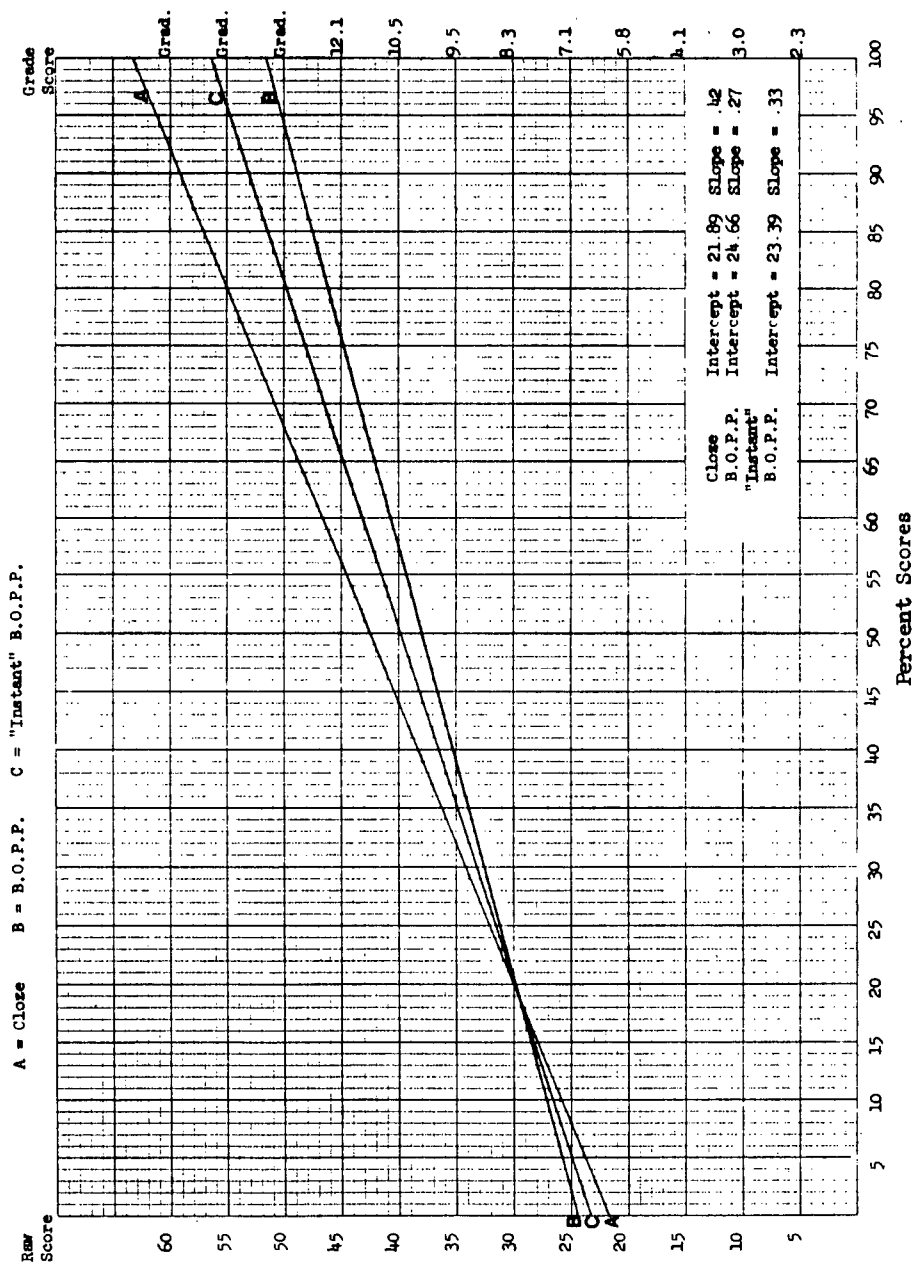
CHAPTER IV

ANALYSIS OF DATA, SUMMARY, CONCLUSIONS AND IMPLICATIONS, RECOMMENDATIONS FOR FURTHER STUDY.

Research Questions

1. Will the cloze procedure, the Beginning of the Page Procedure and the "instant" Beginning of the Page Procedure be positively correlated with the Stanford Diagnostic ?

2. Are the scores yielded by the cloze procedure, the Beginning of the Page Procedure and the "instant" Beginning of the Page Procedure equivalent?

3. What is the difference between the Fry estimate of readability for the passage and the Stanford Diagnostic grade equivalent for 40 percent on the cloze procedure?

4. Is there a significant difference between the performance levels of males and females on the Stanford Diagnostic, the cloze procedure, the

B.O.P.P. or the "instant" B.O.P.P.?

Tests of Research Questions

The answer to question one was found to be positive for all three groups. When the percent scores on the cloze procedure the B.O.P.P. and the "instant" B.O.P.P. were correlated with the raw scores on the Stanford Diagnostic, respective correlations of .54, .53, and .67 were found. (Table X)

The answer to question two was found to be false. A score of 40 percent on the cloze procedure was found to be approximately equivalent to 50 percent on the B.O.P.P. and 45 percent on the "instant" B.O.P.P. (Table VIII)

In answer to question three, a difference was found between the two estimates of readability. The Fry Readability Graph estimated the passage to be at the grade 7.5 level. The grade score on the Stanford Diagnostic , estimated to be equivalent to 40 percent on the cloze procedure, was found to be 10.1 (when comparisons were made with raw scores which were then converted to grade equivalents). Grade equivalents for the B.O.P.P. and "instant" B.O.P.P. were found to be 10.1 and 10.1 respectively. No difference could be expected between these scores as the grade score equivalents for the B.O.P.P. and "instant"

B.O.P.P. scores were obtained through a comparison with cloze procedure scores. (Figure XI, Table XIII)

The answer to question four, was found to be negative. (Tables VI - VII)

TABLE I

MEAN AND STANDARD DEVIATION OF <u>STANFORD DIAGNOSTIC</u> SCORES FOR GROUPS CLOZE PROCEDURE, B.O.P.P. AND "INSTANT" B.O.P.P.

GROUP	MEAN	STANDARD DEVIATION
CLOZE PROCEDURE	38.25	9.97
B.O.P.P.	39.19	10.39
"INSTANT" B.O.P.P.	39.71	10.08

TABLE II

MEAN AND STANDARD DEVIATION OF <u>STANFORD DIAGNOSTIC</u> SCORES FOR MALE AND FEMALE POPULATIONS.

SEX	MEAN	STANDARD DEVIATION
MALE	39.30	10.18
FEMALE	38.77	10.11

TABLE III

ANOVA -- EFFECTS OF SEX ON STANFORD DIAGNOSTIC MEAN
SCORES FOR THE TOTAL POPULATION AND GROUPS CLOZE
PROCEDURE, B.O.P.P. AND "INSTANT" B.O.P.P.

TEST	MEAN SQUARE	F-STATISTIC	SIGNIFICANCE
CLOZE PROCEDURE	.31921	.31576	.9554
B.O.P.P.	.40275	.36716	.9519
"INSTANT" B.O.P.P.	25.615	.24873	.6197
STANFORD DIAGNOSTIC	13.93	.13546	.7132

Significance level = .05

No significant sex differences were found at the .05
level.

TABLE IV

DISTRIBUTION OF SUBJECTS BY SEX

GROUP	MALE	FEMALE	TOTAL
STANFORD DIAGNOSTIC	106	90	196
CLOZE PROCEDURE	35	29	64
B.O.P.P.	33	34	67
"INSTANT" B.O.P.P.	38	27	65

TABLE V

MEAN AND STANDARD DEVIATION FOR PERCENT SCORES FOR
GROUPS CLOZE PROCEDURE, B.O.P.P. AND "INSTANT" B.O.P.P.

GROUP	MEAN	STANDARD DEVIATION
CLOZE PROCEDURE	39.391	12.966
B.O.P.P.	53.567	20.459
"INSTANT" B.O.P.P.	49.477	20.508

TABLE VI

CELL AND MARGINAL MEANS FOR THE STANFORD DIAGNOSTIC RAW
SCORES, AND CLOZE PROCEDURE, B.O.P.P. AND "INSTANT"
B.O.P.P. PERCENT SCORES

TEST	MALE	FEMALE	MARGINAL MEAN
STANFORD DIAGNOSTIC	39.30	38.77	39.06
CLOZE PROCEDURE	40.45	38.51	39.39
B.O.P.P.	56.70	50.53	53.56
"INSTANT" B.O.P.P.	51.03	47.30	49.48

TABLE VII

SUMMARY OF ANOVA EFFECTS OF SEX ON STANFORD DIAGNOSTIC RAW SCORES, AND CLOZE PROCEDURE , B.O.P.P. AND "INSTANT" B.O.P.P. PERCENT SCORES

TEST	MEAN SQUARES	F-STATISTIC	SIGNIFICANCE
STANFORD DIAGNOSTIC	13.94	.13546	.7132
CLOZE PROCEDURE	59.319	.34755	.5576
B.O.P.P.	637.01	1.5341	.2199
"INSTANT" B.O.P.P.	219.61	.51825	.4743

Significance level = .05

No significant sex differences were found at the .05 level.

TABLE VIII

ESTIMATED INSTRUCTIONAL RANGE EQUIVALENCIES FOR GROUPS
CLOZE PROCEDURE, B.O.P.P. AND "INSTANT" B.O.P.P.

GROUP	LOWER LIMIT	UPPER LIMIT	LOWER LIMIT GRADE EQUIVALENT
CLOZE PROCEDURE	40%	59%	10.1%
B.O.P.P.	50%	80%	10.1%
"INSTANT" B.O.P.P.	45%	68%	10.1%

TABLE IX

ESTIMATED EQUIVALENCY TABLE FOR CLOZE PROCEDURE, B.O.P.P. AND "INSTANT" B.O.P.P. SCORES AS PREDICTED FROM STANFORD DIAGNOSTIC RAW SCORES.

STANFORD DIAGNOSTIC RAW SCORE	STANFORD DIAGNOSTIC GRADE SCORE	CLOZE PROCEDURE PERCENT SCORE	B.O.P.P. PERCENT SCORE	"INSTANT" B.O.P.P. PERCENT SCORE
25	7.1	8	2	5
30	8.3	20	20	20
35	9.5	32	40	35
40	10.5	44	57	50
45	12.1	56	75	65
50	GRADUATE	68	94	80
55	GRADUATE	80	—	96
60	GRADUATE	92	—	—

TABLE X

INTERCORRELATIONS OF VARIABLES

VARIABLE	STANFORD DIAGNOSTIC	CLOZE PROCEDURE	B.O.P.P.	"INSTANT" B.O.P.P.
STANFORD DIAGNOSTIC	1.00	.5413	.5341	.6703

TABLE XI

SIGNIFICANCE OF CORRELATIONS OF ALL VARIABLES

VARIABLE	STANFORD DIAGNOSTIC	CLOZE PROCEDURE	B.O.P.P.	"INSTANT" B.O.P.P.

| STANFORD DIAGNOSTIC | .00 | .00 | .00 | .00 |

Significance level = .05

All correlations are highly significant.

TABLE XII

GRADE EQUIVALENTS CORRESPONDING TO STANFORD DIAGNOSTIC RAW SCORES

TEST 1 Reading Comprehension
Total

Raw Score	Grade Equivalent	Raw Score	Grade Equivalent
60	GRAD	30	8.3
59	GRAD	29	8.1
58	GRAD	28	7.8
57	GRAD	27	7.6
56	GRAD	26	7.4
55	GRAD	25	7.1
54	GRAD	24	6.9
53	GRAD	23	6.6
52	GRAD	22	6.4
51	GRAD	21	6.1
50	GRAD	20	5.8
49	GRAD	19	5.4
48	GRAD	18	5.1
47	GRAD	17	4.7
46	12.7	16	4.4
45	12.1	15	4.1
44	11.7	14	3.8
43	11.3	13	3.5
42	11.0	12	3.3
41	10.7	11	3.1
40	10.5	10	3.0
39	10.3	9	2.8
38	10.1	8	2.7
37	9.9	7	2.6
36	9.7	6	2.4
35	9.5	5	2.3
34	9.3	4	2.2
33	9.0	3	2.1
32	8.8	2	1.9
31	8.6	1	1.8

Summary

This research indicated the cloze procedure, the B.O.P.P. and the "instant" B.O.P.P. were all statistically positively correlated with the Stanford Diagnostic which was used as the anchor test. Of the three tests, the "instant" B.O.P.P. was found to be the most highly correlated with the Stanford Diagnostic.

Although the B.O.P.P. and "instant" B.O.P.P. yielded considerably higher percentage scores than the cloze procedure, the high correlation of all test scores with the anchor test seemed to indicate that many of the skills required to complete the cloze procedure, B.O.P.P. and "instant" B.O.P.P. were also the skills required to complete the Stanford Diagnostic. The higher scores on the B.O.P.P. and "instant" B.O.P.P. indicated that students found these to be easier tasks, likely due to the large number of partial words which provided clues to the total word. However, the high correlation of both the B.O.P.P. and "instant" B.O.P.P. to the Stanford Diagnostic suggested that both were valid measures for assessing readability.

When comparing the readability level of the given passage as estimated by the Fry Readability Graph and the readability of the same passage as estimated by the Stanford Diagnostic grade score equivalent to 40 percent on the cloze procedure, it appeared at first glance that

the readability estimates were quite different. However, this study contended that when the necessary adjustments were made to the readability scores, both estimates of readability were virtually the same. This argument was based on the fact that the Fry Readability Graph was formulated using the 50 to 75 percent criterion on the McCall Crabbs Test Lessons , this being the frustration to instructional level. The cloze procedure, on the other hand, was validated against the McCall Crabbs Test Lessons using the 75 to 90 percent criterion, or the instructional to independent level. Burmeister (1974) stated that the difference between a student's frustration and instructional levels was estimated to be one to two years. It was felt that this one to two years must be added to the Fry Readability score before it could be compared to a cloze procedure score. Further, silent reading tests such as the Stanford Diagnostic , were known to inflate the grade scores to the point where the grade scores yielded were usually indicative of the student's frustration level (Burmeister 1974). Burmeister suggested that we must drop back a year or more to find the instructional level. The situation then existed where one to two years were to be added to the Fry score to indicate the instructional level of the passage and one to two years were to be subtracted from the Stanford Diagnostic grade

score equivalent to 40 percent on the cloze procedure (this study has settled on an adjustment score of 1 1/2 years). When these calculations were complete it was found that the Fry Graph indicated the passage to be 7.5 + 1.5 = 9.0 while the _Stanford_Diagnostic_ equivalent of the 40 percent cloze procedure estimated the passage to be 10.1 - 1.5 = 8.6 and thus both gave relatively equivalent estimates of the passages readability.

There was no significant difference found between male and female achievement on the _Stanford_Diagnostic_ results for the entire population, the _Stanford_ _Diagnostic_ results for any of the three groups, cloze procedure, B.O.P.P. and "instant" B.O.P.P. or for the percent scores for these same three groups.

Discussion

The results of this study indicated that the B.O.P.P. and the "instant" B.O.P.P. were appropriate readability measures. The study also showed that when the necessary calculations had been made to both the Fry readability score and the _Stanford_Diagnostic_ grade score equivalent to 40 percent, then both measures estimated the readability level of the passage to be relatively the same. The study, however, was limited in

that only one passage was tested.

Several problems were experienced in using the Stanford Diagnostic which differentiated between scores at the lower end of the scale much more efficiently than between those at the top end of the scale. Grade equivalents were given only to the end of grade twelve and all higher scores were designated as "grad". Unfortunately , 28 percent of the population fell within this category. In the first attempts to analyse the data all calculations were made using grade scores and anyone scoring above the grade 12.9 level was arbitrarily assigned the level 13.5 . This procedure resulted in strong ceiling effects and it appeared advisable to recalculate the data using raw scores which would at least give a distribution of scores at the top end, (if not the grade score equivalent). A test that differentiated well between scores at both ends of the scale would certainly have been a preferable instrument, as the distribution would have been less likely to be skewed in either direction.

No sex differences of any significance were found for any of the groups but contrary to what is usually expected, the males scored higher, although not significantly so, for all categories.

Conclusions and Implications.

The correlation of B.O.P.P. and "instant" B.O.P.P. percent scores with the raw score on the Stanford Diagnostic was shown to be similar to, or higher, than the correlation of the cloze procedure to the Stanford Diagnostic. The cloze procedure had long been recognized as a valid measure of readability (Review of the Literature, Chapter 2) and the results of this study indicated that the B.O.P.P. and "instant" B.O.P.P. were also valid measures of readability. It is true that the percent scores on both these tests tended to be higher than those for the cloze procedure. This was very likely the result of the clues offered by the many partial words, but rather than being a criticism of the B.O.P.P. and "instant" B.O.P.P., Boyce (1974), considering a similar situation, suggested that it gave the student all the contextual clues available in regular reading.

The B.O.P.P. and "instant" B.O.P.P. hold great promise; they have the advantage of the cloze procedure in that they measure the student's ability to deal directly with the material at hand but the time required to prepare and administer these tests, particularly the "instant" B.O.P.P., is considerably less than that required for the cloze procedure. This decrease in time is very important as it increases the likelihood that

such a measure will be used by the practitioner. These conclusions, however, are based on studies involving only one passage and generalizations made from such a study must be questioned. Certainly the percentage scores establishing the instructional level for both the B.O.P.P. and "instant" B.O.P.P. should not be generalized to other materials without further validation.

For the passage studied, the Fry Graph and Stanford Diagnostic grade score equivalent to 40 percent on the cloze procedure, appeared to yield almost the same readability scores once previously mentioned adjustments were made. Once again the study was not broad enough in scope to allow this information to be generalized to other passages.

Recommendations for Future Study

1. Since the study showed promising correlation between the Stanford Diagnostic and both the B.O.P.P. and "instant" B.O.P.P., the study should be replicated using a variety of passages and further, several different grade levels should be involved in the new study.

2.   Various cloze procedure forms for each passage selected should be sampled in order to ensure that the passage chosen is representative .

3.   The strip of paper used to make the B.O.P.P. and "instant" B.O.P.P. should be placed in the middle of the page and at the right hand side of the page. The resulting passages could then be administered to a sample population to ensure that the B.O.P.P. and "instant" B.O.P.P. are no more or less difficult than tests created by using middle or end of the page deletions.

4.   Since the <u>Stanford Diagnostic</u> does not differentiate well between scores at the top end of the scale and since a large percentage of scores fell within this range it is recommended that a new anchor test be employed.

5.   More research is required to determine if the Fry readability score plus 1.5 years is equal to the grade score on a new anchor test minus 1.5 years. This would have to be established over several passages and with several standardized tests before the validity of such a proposal could be verified.

BIBLIOGRAPHY

Anderson, J. Research on comprehension in reading. I. D. Bracken and E. Malmquist(eds), Improving reading ability around the world. Proceedings of the International Reading Association , 1971, 115 - 121.

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

Asher, S. R. and others. Children's comprehension of high and low - interest material and a comparison of two cloze scoring methods, technical report No. 17., 1976 (ERIC Document Reproduction Service ED 134939) (Abstract)

Beil, D. The emperor's new cloze. Journal of Reading,
    1977, 20, 7, 601 - 604.

Betts, E. A. Foundations of reading instruction. New
    York: American Book Company, 1954.

Bormuth, J. R. Cloze as a measure of readability. IRA
    Conference Proceedings, 1963, 8 , 131 - 134.

Bormuth, J. R. Experimental applications of cloze
    tests. International Reading Association
    Conference, 1964, 9 , 303 - 306.

Bormuth, J. R. Comparisons among cloze test scoring
    methods. Proceedings of the International Reading
    Association, 1965, 10 , 283 - 286.

Bormuth, J. R.   Readability: a new approach. <u>Reading</u>
<u>Research Quarterly,</u> 1966, <u>1</u> , 79 - 132.

Bormuth, J. R.   Comparable cloze and multiple-choice
comprehension test scores.   <u>Journal of Reading,</u>
1967, <u>10</u> , 5 , 291 - 299.

Bormuth, J. R.   Cloze test readability: criterion
reference scores,   <u>Journal of Educational</u>
<u>Measurement,</u> 1968, <u>5</u> , 3 , 189 - 196.

Bormuth, J. R.   The cloze readability procedure.
<u>Elementary English,</u> 1968b , <u>45</u> , 429 - 436.

Bormuth, J. R.   Cloze tests and reading comprehension.
<u>Reading Research Quarterly,</u> 1969, <u>4</u> , 358 - 365.

Bormuth, J. R. Development of readability analysis, final Report. University of Chicago, 1969b . (ERIC Document Reproduction Service No. ED 029166)

Bortnick, R., & Lopardo, G. A Case for cloze in the classroom. 1974. (ERIC. Document Reproduction Service No. ED 094327)

Bortnick, R., & Lopardo, G. The cloze procedure: a multi-purpose classroom tool. Reading Improvement, 1976, 13 , 113 - 117.

Boyce, M. W. Some difficulties in using cloze procedures to assess readability. 1974. (ERIC. Document Reproduction Service No. ED 110921)

Boyce, M. W. Some comment on the use of the cloze for classroom mathematics materials. School Science and Mathematics, 1978, 78 , 1, 9 - 12.

Burmeister, L. E. Reading strategies for secondary school teachers. Menlo Park, California: Addison - Wesley Publishing Co., 1974.

Carver, R. P. Measuring the primary effect of reading: reading storage technique, understanding judgements, and cloze. Journal of Educational Behaviour , 1974, 6 , 249 - 256.

Cohen, J. H. The effect of content area material on cloze test performance. Journal of Reading , 1976, 19 , 247.

Culhane, F. G. Cloze procedure and comprehension. Reading Teacher , Feb. 1970, 23 , 410 - 413.

Cunningham, J. W., & Cunningham, P. M. Validating a limited-cloze procedure. Journal of Reading Behaviour, 1978, 10 , 211 - 213.

Daines, D., & Mason, L.   A comparison of placement   test
    readability graphs.   Journal of Reading , 1972, 15 ,
    597.

Dale, E., & Chall, J. S.   A Formula for predicting
    readability.   Educational Research Bulletin , 1948,
    48 , 11 - 20 and 28.

Dishner, E. K.   The cloze procedure: a valuable tool for
    content teachers.   1973   (ERIC   Document
    Reproduction Service No.   ED 123602)

Entin, E. B., & Klare, G. R.   Some inter-relationships
    of readability, cloze and multiple-choice scores on
    a reading comprehension test.   Journal of Reading
    Behaviour, 1978, 10 , 4 , 417 - 434.

Farris, L. S.   A comparison of cloze and multiple-choice
    procedures for measuring reading comprehension.
    1975.   (ERIC   Document Reproduction Service No.
    Document ED 120666) (Abstract)

Fry, E. B.   A readability formula that saves time.
    Journal of Reading, April 1968, 11 , 513 - 516,
    575 - 578.

Fry, E. B.   Fry's readability graph: clarifications,
    validity and extension to level 17, Journal of
    Reading, Dec. 1979, 21 , 3 , 242 - 252.

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

Hafner, L. E.   Relationships of various measures to the
    "cloze". National Reading Conference , 1964, 13 ,
    137 - 145.

Hafner, L. E.   Implications of cloze. National Reading
    Conference , 1965, 14 , 151 - 158.

Hansell, T. S. Readability, syntactic transformations, and generative semantics. Journal of Reading, 1976, 19 , 557.

Jefferson, G. L. Jr. Lexical and structural items as predictors of readability for high and low ability readers. Nineteenth Yearbook of the National Reading Conference , 1971, 172 - 178.

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

Jongsma, E. R. The cloze procedure: a survey of the research. 1971. (ERIC. Document Reproduction Service No. ED 050893)

Keonk, K. Another practical note on readability formulas. Journal of Reading, 1971, 15 , 203.

Kirby, C. <u>Using the cloze procedure as a testing technique.</u> April 1968. (ERIC. Document Reproduction Service No. ED 019202)

Klare, G. R. <u>The measurement of readability.</u> Ames, Iowa: Iowa State University Press, 1963.

Klare, G. R. Assessing readability. <u>Reading Research Quarterly,</u> 1975, <u>10</u>, 62 - 100.

Klare, G. R. A second look at the validity of readability formulas. <u>Journal of Reading Behaviour,</u> 1976, <u>8</u>, 129 - 152.

McCabe, P. P. Give readers a B.O.P.P. <u>Journal of Reading,</u> Dec. 1979, <u>23</u>, 3, 199.

MacGinitie, W. H. Contextual constraint in english prose paragraphs. Journal of Psychology, 1961, 51 .

McKenna, M. Synonymic versus verbatim scoring of the cloze procedure. Journal of Reading , 1976, 20 , 2, 141 - 143.

McLaughlin, G. H. Smog grading - a new readability formula. Journal of Reading, 1969, 12 , 639.

Miller, L. R. Predictive powers of multiple-choice and cloze derived readability formulas. Reading Improvement , Spring 1975, 12 , 1 , 52 - 58.

Myers, P. C. The cloze procedure: latest research and uses. 1976. (ERIC. Document Reproduction Service No. ED 123556)

Oller, J. W. Jr. Cloze: discourse and approximations to English. 1975. (ERIC. Document Reproduction Service No. ED 107144) (Abstract)

Panackal, A. A., & Heft, C. S. Cloze technique and multiple-choice technique: reliability and validity. Educational and Psychological Measurement, 1978, 38 , 4 , 917 - 932.

Pauk, W. A practical note on readability formulas. Journal of Reading, 1969, 13 , 207.

Pennock, C. D. Selecting instructional materials. Alberta Journal of Educational Research, 1973, 19 , 1 , 30 - 36.

Pennock, C. D. The cloze test for assessing readability. English Quarterly, 1973, 5 , 4 , 35 - 39.

Peterson, J. and others. Validation of the cloze procedure as a measure of readability with high school, trade school and college populations. 1972. (ERIC. Document Reproduction Service No. ED 058000)

Pollock, D. H. The use of cloze to determine reading instructional levels. (ERIC. Document Reproduction Service No. ED 099792) (Abstract)

Potter, T. C. A taxonomy of cloze research part I: readability and reading comprehension. 1968. (ERIC. Document Reproduction Service No. ED 022644) (Abstract)

Rakes, T. A., & McWilliams, L. Bridging the gap: two alternatives to standardized testing. English Journal, Oct. 1978, 67, 7, 46 - 50.

Rankin, E. F. Jr. The cloze procedure - its validity and utility. Eighth Yearbook of the National Reading Conference, 1959, 8 , 131 - 144.

Rankin, E. F. Jr. The cloze procedure - a survey of research. National Reading Conference, 1965, 14 .

Rankin, E. F. Grade level interpretations of cloze readability scores. Dec. 1970. (ERIC. Document Reproduction Service No. ED 046657)

Rankin, E. F., & Culhane, J. W. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1969, 13 , 3 , 193 - 198.

Riley, P. M. The cloze procedure -- a selected annotated bibliography. 1973. (ERIC. Document Reproduction Service No. ED 106749)

Rosenkranz, C. I. R. The efficiency of cloze procedure for estimating reading ability of students and readability of materials in adult fundamental educational programs. (ERIC. Document Reproduction Service No. ED 119127) (Abstract)

Ruddell, R. B. A study of the cloze comprehension technique in relation to structurally controlled reading material. International Reading Association Conference, 1964, 9 298 - 303.

Rupley, W. H. The cloze procedure. Journal of Reading, 1973, 16 , 6 , 496 - 497.

Russell, S. N. How effective is the cloze technique in the measurement of reading comprehension? Viewpoints in Teaching and Learning, 1978, 54 , 3 , 90 - 95.

Schneyer, J. W. Use of the cloze procedure for improving reading comprehension. Reading Teacher, 1965, 19 , 174 - 179.

Schoelles, I. S. Cloze as a predictor of reading group placement. April, 1971. (ERIC. Document Reproduction Service No. ED 053868)

Smith, N., & Zink, A. A cloze-based investigation of reading comprehension as a composite of subskills. Journal of Reading Behaviour, 1977, 9 , 4 , 395 - 398.

Spooncer, F. Hanging out the cloze line. United Kingdom Reading Association, June 1974, 8 , 19 - 26.

Taylor, W. R. "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.

Taylor, W. L. "Cloze" readability scores as indices of
    individual   differences   in   comprehension   and
    aptitude. *Journal of Applied Psychology,* 1957, 41 ,
    1 , 19 - 26.

Thelen, J. N.   Using the cloze test with science
    textbooks. *Science and Children,* 1974, 12 , 3 , 26
    - 27.

Thorndike, E. L. Reading and reasoning: a study of
    mistakes   in   paragraph   reading.   *Journal of
    Educational Psychology,* 1917, 8 , 323 - 332.

Tinzmann, M. B., & Thomas, G. R.   *A comparison of
    listening and reading cloze procedures and a
    standardized reading achievement test.*   May, 1977.
    (ERIC. Document Reproduction Service   No.   ED
    162249)

Tuinman, J., & Blanton, G.   A note on cloze as a measure
of comprehension.   Journal of Psychology, 1975, 90 ,
159 - 162.

Van   Rooy,   L.   Readability studies and the writer of
instructional materials.   1973.   (ERIC.   Document
Reproduction Service No.   ED 089245)

Vaughan,   J. L. Jr.   Interpreting   readability
assessments.   Journal of Reading, 1976, 19 , 635.

Walter,   R. B.   Historical overview of the cloze
procedure.   1974.   (ERIC.   Document Reproduction
Service No.   ED 094337)

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

Weaver, G. C.   Using the cloze procedure as a teaching
     technique. Reading Teacher, 1979, 32 , 5 ,   632   -
     636.

Weaver, W. W.   Theoretical   aspects   of   the   cloze
     procedure. National Reading Conference , 1964, 13 ,
     115 - 132.

Appendix A

Corrected Lorge Formula

Compute average sentence length in words (X2);

Compute number of prepositional phrases per 100 words (X3);

Count number of <u>different</u> hard words not on the Dale 769 word list (X4);

Substitute in the formula:

$$X1(grade\ placement) = .06X2 + .10X3 + .10X4 + 1.99$$

X1 stands for the average reading ability required to correctly answer one-half of the test questions on a given passage. (Klare, 1963)

Appendix B

Flesch Corrected Formula

The Flesch formula, unlike the Lorge, was not designed to give a reading grade level but rather to indicate a level of difficulty based on seven classes, 1 being the easiest and 7 being the most difficult.

Systematically select samples of 100 words throughout the material to be rated:

Compute average sentence length in words (Xs);

Count the number of affixes (Xm);

Count the number of personal references (Xh);

Average the results and insert in the formula:

$$.07Xm + .07Xs - .05Xh + 3.27 \text{ (Klare 1963)}$$

Flesch stipulated that the users of his formula were to count as sentences each unit of thought that was gramatically independent of another sentence or clause, if it ended with a period, question mark, exclamation point, semicolon or colon. Sentence fragments were also to be counted as sentences.

Appendix C

Flesch Reading Ease Formula

Systematically select 100-word samples from material
to be rated;

Determine the number of syllables per 100 words
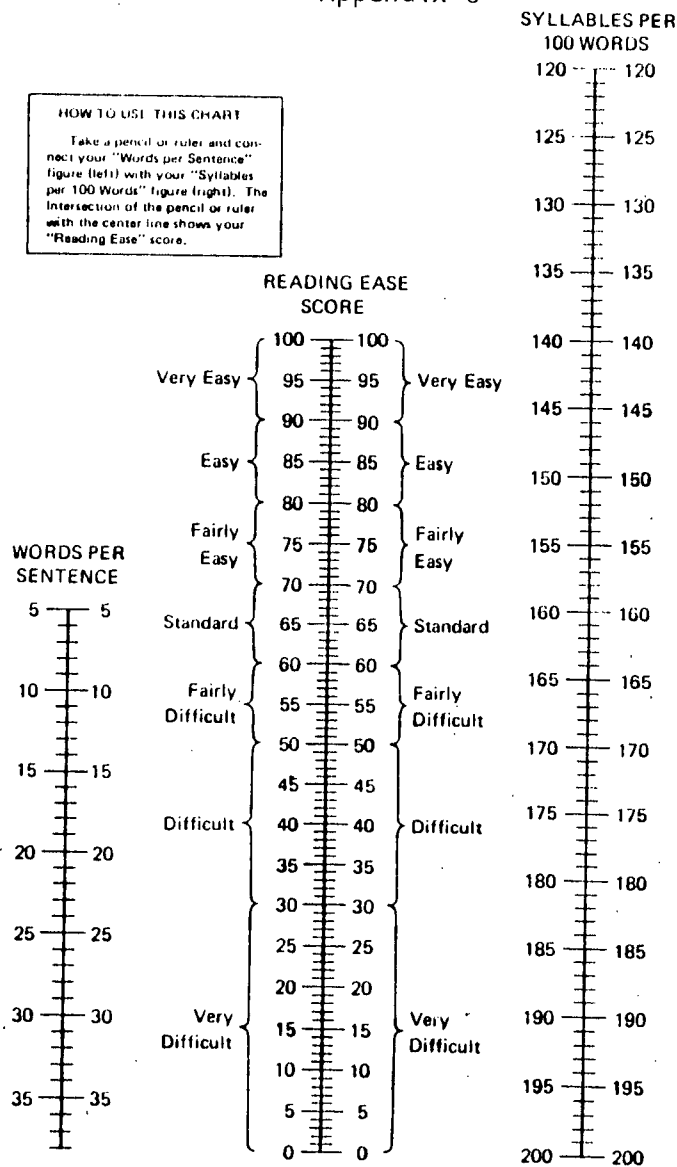(WL);

Determine the average number of words per sentence
(SL);

Apply in the following reading ease equation:

$$R.E. = 206.835 - .846WL - 1.01SL \quad \text{(Klare 1963)}$$

Refer to charts for level of difficulty and approximate
grade equivalent.

Appendix C

SYLLABLES PER 100 WORDS

120 — 120
125 — 125
130 — 130
135 — 135
140 — 140
145 — 145
150 — 150
155 — 155
160 — 160
165 — 165
170 — 170
175 — 175
180 — 180
185 — 185
190 — 190
195 — 195
200 — 200

READING EASE SCORE

Very Easy { 100 — 100
95 — 95 } Very Easy
90 — 90

Easy { 85 — 85 } Easy
80 — 80

Fairly Easy { 75 — 75 } Fairly Easy
70 — 70

Standard { 65 — 65 } Standard
60 — 60

Fairly Difficult { 55 — 55 } Fairly Difficult
50 — 50
45 — 45

Difficult { 40 — 40 } Difficult
35 — 35
30 — 30
25 — 25
20 — 20

Very Difficult { 15 — 15 } Very Difficult
10 — 10
5 — 5
0 — 0

WORDS PER SENTENCE

5 — 5
10 — 10
15 — 15
20 — 20
25 — 25
30 — 30
35 — 35

© 1949 by Rudolf Flesch

Reading-ease score	Description of style	Typical magazine	Grade
90 to 100	Very easy	Comics	5
80 to 90	Easy	Pulp fiction	6
70 to 80	Fairly easy	Slick fiction	7
60 to 70	Standard	Digests, *Time*, Mass nonfiction	8 and 9
50 to 60	Fairly difficult	*Harper's, Atlantic*	10–12 (high school)
30 to 50	Difficult	Academic, scholarly	13–16 (college)
0 to 30	Very difficult	Scientific, professional	College graduate

Flesch, R. The art of readable writing. New York: Harper and Row Publishing Co., 1949, p.5.

Flesch, R. How to test readability. New York: Harper and Row Publishing Co., 1951, pp. 6, 43.

Appendix D

Dale-chall Formula

Select 100-word samples throughout the material to
be rated;

(about every tenth page for books, every 2000 words
for articles)

Compute the average sentence length in words (X2);

Compute the percentage of words outside the Dale
list of 3000 (X1);

Apply in formula:

$$Xc50 = .1579x1 + .0496x2 + 3.6365$$

Where Xc50 refers to the reading grade score of a
student who can answer one-half of the test questions on
a passage correctly. (Klare 1963) Dale and Chall (1948)
set up the following table of estimated corrected grade
levels:

Appendix D

Formula Score	Corrected Grade level
4.9 and below	Grade IV and below
5.0 to 5.9	Grades V - VI
6.0 to 6.9	Grades VII - VIII
7.0 to 7.9	Grades IX - X
8.0 to 8.9	Grades XI - XII
9.0 to 9 .9	Grades XIII - XV (College)
10.00 and above	Grades XVI +College graduate

The Dale-Chall list has not been included.

Appendix E

GRAPH FOR ESTIMATING READABILITY —EXTENDED

by Edward Fry. Rutgers University Reading Center, New Brunswick, N.J 08904

Average number of syllables per 100 words
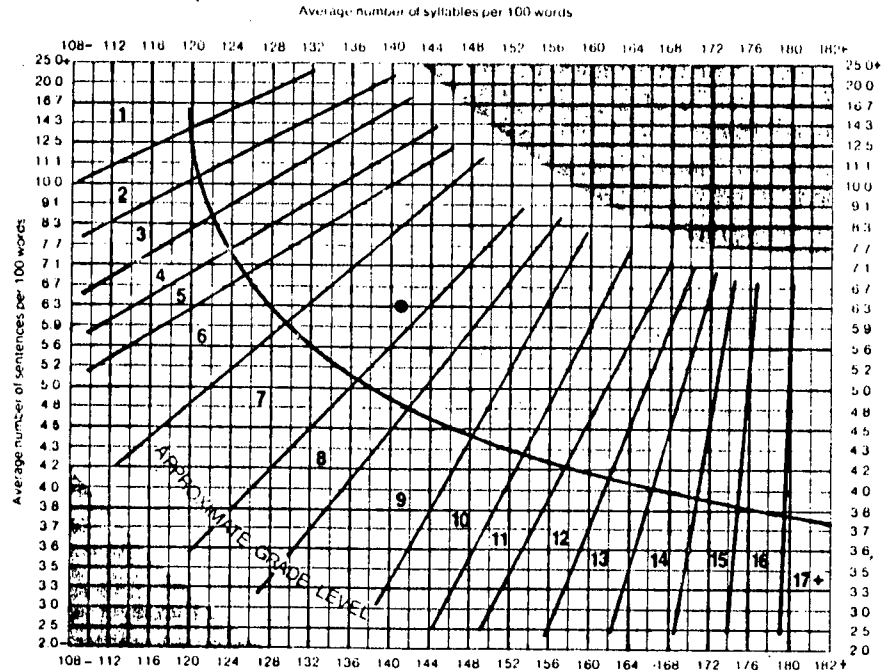
Expanded Directions for Working Readability Graph

1. Randomly select three (3) sample passages and count out exactly 100 words each, beginning with the beginning of a sentence. Do count proper nouns, initializations, and numerals.
2. Count the number of sentences in the hundred words, estimating length of the fraction of the last sentence to the nearest one-tenth.
3. Count the total number of syllables in the 100-word passage. If you don't have a hand counter available, an easy way is to simply put a mark above every syllable over one in each word, then when you get to the end of the passage, count the number of marks and add 100. Small calculators can also be used as counters by pushing numeral 1, then push the + sign for each word or syllable when counting.
4. Enter graph with *average* sentence length and *average* number of syllables; plot dot where the two lines intersect. Area where dot is plotted will give you the approximate grade level.
5. If a great deal of variability is found in syllable count or sentence count, putting more samples into the average is desirable.
6. A word is defined as a group of symbols with a space on either side; thus, *Joe, IRA, 1945,* and & are each one word.
7. A syllable is defined as a phonetic syllable. Generally, there are as many syllables as vowel sounds. For example, *stopped* is one syllable and *wanted* is two syllables. When counting syllables for numerals and initializations, count one syllable for each symbol. For example, *1945* is four syllables, *IRA* is three syllables, and & is one syllable.

Note: This "extended graph" does not outmode or render the earlier (1968) version inoperative or inaccurate; it is an extension. (REPRODUCTION PERMITTED—NO COPYRIGHT)

Appendix F

Smog Readability Formula

SMOG grading = 3+ square root of polysyllable count. The polysyllable count is the number of words, within a thirty sentence passage, that have three or more syllables.

SMOG Grading

1. Count 10 consecutive sentences near the beginning of the text to be assessed, 10 in the middle and 10 near the end. Count as a sentence any string of words ending with a period, question mark or exclamation point.

2. In the 30 selected sentences count every word of three or more syllables. Any string of letters or numerals beginning and ending with a space or punctuation mark should be counted if at least three syllables can be distinguished when it is read aloud in context . If a polysyllabic

word is repeated, count each repetition.

3.   Estimate the square root of the number of
       polysyllabic words counted. This is done
       by taking the square root of the nearest
       perfect square. For example, if the count
       is 95, the nearest perfect square is 100,
       which yields a square root of 10. If the
       count lies roughly between two perfect
       squares, choose the lower number. For
       instance, if the count is 110, take the
       square root of 100 rather than that of
       121.

4. Add 3 to the approximate square root.
    This gives the SMOG Grade, which is the
    reading grade that a person must have
    reached if he is to understand fully the
    text assessed.

Appendix G

THE GOOD OLD DAYS?

Have you ever been told, "Now you're going to catch it! Just wait till your _____ gets home," or "Wait _____ your mother finds out"? _____ you had lived in _____ times, you might have _____ even more apprehensive if _____ had been told, "Wait _____ Lecture Day!"

Colonial punishment _____ misconduct could be quite _____. To be embarrassed and _____ by the whole town _____ one of the agonizing _____ endured by many.

On _____ Day, all the community _____ aside its work, packed _____ and went to the _____ square. There, a preacher _____ deliver a lengthy lecture - _____ dramatic fire-and-brimstone _____ on the consequences of _____ behavior. It was designed _____ put fear of misbehavior _____ listeners' hearts. Everyone listened _____; but it was the _____ that the villagers awaited.

_____ the speech was finally _____, all those convicted of _____ were paraded to a _____ in front of the _____. They were forced to _____ their guilt and publicly _____. Then they were whipped.

_____ real criminals - those who _____ murdered or robbed large _____ - were hanged, as were _____ accused of witchcraft. Others - _____ thieves, for instance - were _____. The rest were locked _____ the stocks or pillory. _____ were those accused of _____ beating, cursing, nagging, drunkenness, _____ to observe the Sabbath, _____ talking back to parents.

_____ stocks were a wooden _____ which restrained a seated _____ by fastening hands and _____ in locked frames. The _____ restrained a person's head _____ hands. The punishment was _____ to be psychological, but _____ passersby added physical torment _____ pelting prisoners with stones.

The idea of public ridicule was a terrible one, and was effective in keeping most people within the binding rules.

Appendix H

till your fa	1	:ts home," or "Wait till your mother finds out"? If you	1._____
had lived in	2	al times, you might have been even more apprehensive if you	2._____
had been tol	3	t until Lecture Day!"	3._____

Colonia	4	hment for misconduct could be quite severe. To be	4._____
embarrassed	5	graced by the whole town was one of the agonizing punishments	5._____
endured by m.	6		6._____

On Lect	7	', all the community put aside its work, packed lunches and	7._____
went to the	8	juare. There, a preacher would deliver a lengthy lecture - a	8._____
dramatic fir	9	rimstone sermon on the consequences of bad behavior. It was	9._____
designed to	10	ir of misbehavior into listeners' hearts. Everyone listened	10._____
quietly; but	11	the follow-up that the villagers awaited.	11._____

When th	12	:h was finally over, all those convicted of crimes were	12._____
paraded to a	13	rm in front of the people. They were forced to admit their	13._____
guilt and pu	14	apologize. Then there were whipped.	14._____

The rea	15	inals - those who had murdered or robbed large amounts- were	15._____
hanged, as w	16	se accused of witchcraft. Others - common thieves, for	16._____
instance - w	17	risoned. The rest were locked in the stocks or pillory. They	17._____
were those a	18	of wife beating, cursing, nagging, drunkenness, failure to	18._____
observe the	19	, or talking back to parents.	19._____

The sto	20	e a wooden structure which restrained a seated prisoner by	20._____
fastening ha	21	feet in locked frames. The pillory restrained a person's	21._____
head and han	22	e punishment was meant to be psychological, but often passersby	22._____
added physic	23	ent by pelting prisoners with stones.	23._____

The idea of public ridicule was a terrible one, and was effective in keeping most people within the binding rules.

Appendix I

Unit 19 — THE GOOD OLD DAYS?

Have you ever been told, "Now you're going to catch it! Just wait
till your father   1.   ›me," or "Wait till your mother finds out"? If you
had lived in col   2.   mes, you might have been even more apprehensive
if you had been   3.   Wait until Lecture Day!"

1. _____
2. _____
3. _____

Colonial p   4.   ıent for misconduct could be quite severe. To be
embarrassed aı   5.   raced by the whole town was one of the agonizing
punishments eı   6.   by many.

4. _____
5. _____
6. _____

On Lectuı   7.   all the community put aside its work, packed
lunches and wɛ   8.   he town square. There, a preacher would deliver a
lengthy lecturı   9.   amatic fire-and-brimstone sermon on the conse-
quences of bad   10.   or. It was designed to put fear of misbehavior into
listeners' heart   11.   yone listened quietly; but it was the follow-up that
the villagers aw   12.

7. _____
8. _____
9. _____
10. _____
11. _____
12. _____

When the   13   was finally over, all those convicted of crimes were
paraded to a p   14.   in front of the people. They were forced to admit
their guilt and   15.   apologize. Then they were whipped.

13. _____
14. _____
15. _____

The real   16.   ıls—those who had murdered or robbed large
amounts—wer   17.   ɛd, as were those accused of witchcraft. Others
—common thiɛ   18.   instance—were imprisoned. The rest were locked
in the stocks o   19.   y. They were those accused of wife beating, curs-
ing, nagging,   20.   ıness, failure to observe the Sabbath, or talking
back to parentɛ

16. _____
17. _____
18. _____
19. _____
20. _____

The stocł   21.   a wooden structure which restrained a seated
prisoner by fɛ   22.   hands and feet in locked frames. The pillory
restrained a pɛ   23.   ıead and hands. The punishment was meant to be
psychological,   24.   ɛn passersby added physical torment by pelting
prisoners with   25.

21. _____
22. _____
23. _____
24. _____
25. _____

The idea of public ridicule was a terrible one, and was effective in
keeping most people within the binding rules.