DOCUMENT RESUME

ED 110 921 CS 001 984

AUTHOR Boyce, Max William
TITLE Some Difficulties in Using Cloze Procedures to Assess Readability.
PUB DATE Apr 74
NOTE 138p.; M.Ed. Thesis, University of Melbourne

EDRS PRICE MF-$0.76 HC-$6.97 Plus Postage
DESCRIPTORS *Cloze Procedure; Elementary Education; Grade 6; *Independent Reading; Instructional Materials; Measurement Instruments; *Readability; Reading Research; *Reading Skills; *Test Construction; Testing

ABSTRACT

This thesis explores some difficulties associated with the use of the cloze procedure, particularly in relation to the interpretation of an individual's score on a cloze test. Cloze tests were administered to 112 grade-six children in four schools. The results indicated that for the children in this study the easiest words to replace were those that are one syllable long, or articles, or conjunctions, or prepositions, or pronouns. It is suggested that there is likely to be considerable overlap between the expected scores for the independent and instructional reading levels. Some of the limitations of the study are discussed. (Author)

SOME DIFFICULTIES IN USING CLOZE PROCEDURES TO ASSESS READABILITY

Max William Boyce

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Education by course work in the University of Melbourne.
April 1974.

TABLE OF CONTENTS

APPENDICES cont'd.

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

In presenting this thesis I wish to record my sincere thanks to the many people who have provided assistance, guidance and encouragement during the project.

My supervisor for the project was Mr. Charles Poole. His help, particularly in relation to the statistical analysis of the data, was of great assistance. His guidance made the investigation a valuable learning experience.

I am grateful to those third year students at the State College of Victoria, Toorak, who assisted with the administration of the tests for the Pilot Study. For the main investigation I am particularly grateful to Mrs. I. Dimkey and Messrs. J. Storey, W. Jones and G. Spalding, who helped considerably by ranking their students, selecting appropriate books, and allowing me to use their grades for the collection of data.

The typing of the thesis was done by Mrs. M. H. Hunter.

Finally I would like to thank those of my colleagues at the State College of Victoria, Toorak, who by their interest and comments gave me great support.

ABSTRACT

This thesis explores some difficulties associated with the use of the cloze procedure, particularly in relation to the interpretation of an individual's score on a cloze test used to determine whether the material from which the test is taken is, or is not, suitable for his instructional or independent reading.

A number of cloze methodological considerations are discussed in detail. The literature relating to the development of comparable cloze and multiple-choice criteria for passage performance is reviewed.

The Pilot Study explores the possibility that any one cloze test from a passage of prose might be more or less difficult than any of the other possible cloze tests from the same passage. After establishing means of categorizing the deleted words, the six possible every sixth word deleted cloze tests from two 300 word passages from two different books at Grade Six reading level were used. On the basis of the categorization of the words deleted in each of these tests, the 'easiest' and 'hardest' tests for each passage were predicted. These were then tested on 196 Grade Six children in eight schools. For both passages the mean score for the predicted 'easiest' passage was significantly higher than that for the predicted 'hardest' passage from the same passage.

The main investigation attempts to establish (i) the characteristics of deleted words which influence the difficulty levels of cloze tests; (ii) a simple means of adjusting the obtained cloze scores to allow for the relative ease or difficulty of replacement of the deleted words; and (iii) an operationally determined range of scores which could serve as a criterion to indicate whether material is suitable for an individual child's independent or unsupervised reading.

For these purposes 112 different 350 word cloze tests were developed (the seven possible every seventh word deleted patterns of 16 different passages). The 5,600 words included in these tests were placed in the categories established in the Pilot Study. The tests were given to 112 Grade Six children in four schools.

The results indicate that for the children in this study the easiest words to replace were those that are one syllable long, and/or are 1-2 letters long, and/or are in common word lists, and/or are articles, conjunctions, prepositions or pronouns. A regression analysis determined two formulas for adjusting the obtained scores but it was decided that the gain achieved would not justify the work involved in using them. The mean and standard deviation gave an estimate of the scores that could be expected to be achieved by two-thirds of the children doing cloze tests from material suitable for their independent reading. It is suggested that there is likely to be considerable overlap between expected scores for the independent and instructional levels.

Some limitations of the study are discussed.

CHAPTER 1

INTRODUCTION

The post-Sputnik era has seen the development of pressures both inside and outside the education profession to improve all education. As a consequence a great deal of public money has been poured into educational research and curriculum development. At the same time there has been a change in the organizational patterns of schools in order to meet the needs of an increasing and divergent school population, as well as to meet the needs of changing philosophies of educational theory and practice.

One major determinant of the changes that have occurred has been the concentration on the individual learner. Although it has been long accepted that individuals are different, and their capacities to learn are different, the gap between acceptance in principle and acceptance in practice took a long time to bridge. It is only in very recent years that any real attempt has been made to individualize school programs. Until recently primary teaching, particularly in the upper part of the school, was generally conceived of in terms of formal instruction with the whole class as the working unit. Curricula were relatively clearly delineated and prescribed, and trainee teachers were taught specific methods to meet the demands of these curricula.

The last decade has seen significant changes. Substantial modifications have been made to the primary school curriculum, leading to a move away from detailed prescription of content to the development of source materials (Warry and Fitzgerald, 1969). These changes have been characterized by a greater emphasis on higher cognitive and affective objectives (Ainley,1972a). Thus, for example, the new mathematics courses emphasize themes and concepts rather than procedural drill, and the new science courses tend to emphasize

process rather than content, with 'discovery learning' playing an important part (See Ainley, 1972b p.29-30).

Concomitant with this change in curricula has been an increasing emphasis on individual and small group instruction, a demand for flexible approaches and for the appropriate conditions for learning (Grace, 1967). In fact in Victoria the traditional classroom with its associated teacher behaviours is no longer officially acceptable - ".... the self contained classroom and the self contained school are obsolete." (Education Gazette and Teachers Aid, 1972, p. 513)

Thus we are now faced with a variety of alternative classroom organizations, ranging from tentative variations on the old traditional theme to open classrooms (of infinite variety, philosophy, definition and effectiveness), family grouping and other multi-grading formats, together with an emphasis upon individualized procedures and increasingly open and flexible curricula.

The changes in classroom organization, the 'opening' of courses and the emphasis on individualization of instruction bring with them a number of problems for the teacher, not the least of which is the need, more than ever before, to ensure that the instructional materials used by the child are suitable for his individual stage of development and ability.

This is, of course, not a new problem. It is one that has always, in theory, existed. It is a problem, however, that has been intensified by modern educational practice, by the needs of the individual and, in some educational systems, by the demands for accountability and responsibility in educational practice. The more we move away from teacher centred, group teaching situations, to independent, individualized learning experiences, the more apparent is the need to ensure that the child can read and effectively cope with the materials provided for him.

There are, therefore, a number of reasons why the teacher needs to have the means of determining the suitability of the material in any given

learning task for the individual. Among these reasons are:

(a) the need to be able to monitor the student's learning during instruction so that instructional procedures and materials can be altered as needed;

(b) the need to provide materials that are difficult enough to challenge but sufficiently easy to ensure success;

(c) the need to decide when a student has gained sufficient mastery of the content to warrant advancing him to a more complex unit;

(d) the need to avoid children being given tasks that are too difficult or too easy and thus running the risk of them unnecessarily wasting time, becoming frustrated or anxious, or developing negative attitudes to self and learning.

Determining suitability of materials.

There are a number of ways in which the suitability of reading material for the individual might be determined.

1. Teacher and librarian estimates.

Generally estimates of the suitability of materials made by teachers and librarians are subjective and as such are often open to a great deal of question. Klare (1963, p.81) states that "they are recognized as subject to considerable error", whilst Russell and Merrill (1951), in a study in which children's librarians rated the difficulty of well known juvenile books, found that such "expert" opinions do not show much general agreement.

2. Readability formulas.

This has been a common method of determining suitability of written materials. For this purpose the term 'readability' refers to the difficulty

level or comprehensibility of written prose.

Readability formulas attempt to predict the likelihood of a given reading selection being understood by an individual or group of individuals. This is done by attempting to label selections of prose in terms of appropriate grade levels. There are a large number of elements involved in the concept of readability and Fry (1972, p.204) makes one of the many attempts to enumerate them. Anderson (1967) and Klare (1963) give analyses of the factors involved.

In general readability formulas make use of regression equations and take into consideration variables such as sentence length, number of syllables and the number of difficult or unfamiliar words.

Ball and Williamson (1973) claim that the formulas devised by Flesch (1943, 1944), Lorge (1944) and Dale and Chall (1948) are ".... simple to apply, yield consistent differentiation of standard sets of passages and have been shown to agree with observations of children's reading performances." (p.14) As a result their Readability levels of Children's Literature (Williamson and Ball, 1973) is based on the use of the Dale-Chall formula.

Whilst readability formulas have been quite widely used by some educationists, librarians, publishers, and others in the field of communication, there are a number of critics of the usefulness of such measures. For example, Blair (1971) believes that there are too many aspects involved in readability that are not included in these formulas, such as contextual difficulty, abstractness of ideas, density of ideas, interest of subject, style appeal, material organization, size of type, type of ink, etc., etc. Otto and Smith (1971) believe that mechanical formulas such as these work in opposition to any concept of readability which accepts the criterion that what an individual can read is, to him, readable. Thus, from their point of view, the only way to determine what is readable is by direct testing on the material by the individual.

Despite Ball and Williamson's (1973) contention that readability formulas are simple to apply, it is probable that their mathematical nature limits their usefulness for many teachers and librarians. For example, the mathematical expressions of the three formulas they mention are as follows:

(a) Flesch Reading Ease = 206.835 - .846 wl - 1.015 sl.

206.835 is a constant.

wl = the number of syllables per 100 words.

sl = the average number of words per sentence.

Reading Ease represents the grade level which would have to be attained in order to read the passage.

(b) Lorge $C_{50} = 0.06_a + 9.55_b + 10.43_c + 1.9892$

1.9892 is a constant

a = average sentence length.

b = ratio of prepositional phrases to total number of words.

c = ratio of hard words (i.e. words not in Dale's 769 'easy words' List) to total number of words.

C_{50} is the reading grade score of a pupil who answers one half of a series of test questions correctly.

(c) Dale and Chall

$C_{50} = .1579_a + .0496_b + 3.6365$

3.6365 is a constant.

a = average sentence length in words.

b = percentage of words outside the Dale list of 3,000 words.

C_{50} is the reading grade score of a pupil who answers one half of a series of test questions correctly.

Even a cursory glance at these formulas would indicate that, even if

one discounts the problem of their mathematical nature, they are tedious and time consuming for the average classroom teacher to use for practical purposes, especially those that require the searching through of lengthy word lists to see if the words in the selected passages are in these lists, and the determination of prepositional phrases.

There is also some contention about whether these formulas yield consistent differentiation. Carozzi (1972) points out that correlations between some of the formulas, e.g. Flesch and Dale and Chall, could be spuriously high as they include a sentence factor in common and have used the same criterion, viz. the McCall-Crabbs Test lessons. Michaelas and Tyler (in Froese, 1971) quote contradictory evidence regarding correlations between such formulas, whilst Blair (1971) maintains that some show consistently higher scores (levels) than others, with the consequence that a readability level depends to a great extent on the measure used. Bormuth (1966) believes that the current formulas may hinder more than help because of their low predictive values and because they make poor guides for adjusting the difficulty of materials.

Finally, although Klare (1952) argues that "... readability formulas are ..... sufficiently accurate for estimating the comparative readability of adult materials" (p.397) (my underlining) and Lorge (1948) points out that the readability index is an estimate and not intended as a precise indication, Carozzi (1972) indicates that "teachers and publishers tend to treat readability formulas as though they were precise measures." (p.71) So although Spache and Chall each pointed out that levels from formulas are only accurate to within $\pm$ 1 year of reading age (McLeod, 1962), the formulas have been used to make distinctions of 1 to 2 months in the reading difficulty of books without also making the error involved quite clear. (See, for example, Bird and Falk, 1971).

3. Direct testing.

An alternative to the subjectivity of teacher/librarian estimates and the problems associated with the use of readability formulas is that of testing the reading material on the child directly.

In most versions of this procedure the student is asked to read a passage that is thought to be representative of the book or instructional materials, and then answer some questions about the passage - the questions usually being of a multiple-choice format.

It has been accepted for a long time (e.g. Kilgallon, 1942; Betts, 1946) that if a child is able to answer at least 90% of the questions based on the material he has read then the material is said to be at his independent level, and is therefore suitable for use in his unsupervised study and voluntary reading. If he is able to answer at least 75% of the questions, then the material is said to be at his instructional level, and suitable for use in his supervised instruction. If he is unable to answer at least 50% of the questions then the materials are said to be too difficult, or unsuitable, or at his frustrational level. These levels, which have been operationally defined, have been used in readability formulas such as those of Dale-Chall, Lorge and Flesch, where the criteria has been based on either 50% or 75% comprehension on the McCall-Crabbs Test lessons.

The direct testing approach has been recommended in a number of reading textbooks, e.g. Bond and Tinker (1967), Della-Piana (1968), Harris (1962) and Russell and Thompson (1966).

There is, however, a major problem associated with the use of direct testing of material on the individual child. That problem is the dependence on multiple-choice questions as the criterion for performance. The problem is accentuated by the fact, that for most practical purposes, it is the teacher himself who writes the multiple-choice questions. Wesman (1971) writes: "Item writing is essentially creative - it is an art.....(it) requires an uncommon combination of special abilities and is mastered only through extensive and critically supervised practice." (p.81) It is probable that very few teachers are sufficiently trained in the skills necessary to construct test questions that can meet the criteria for even relatively loose standards of replicability.

The requirements for items to exhibit the necessary clarity and effectiveness are numerous. Wesman (1971) lists 12 general suggestions as well as another 12 specific suggestions for the writing of multiple-choice items. (See Appendix A)

It is probable therefore that the following difficulties may be associated with direct testing where the criterion is a test involving multiple-choice items devised by the classroom teacher:

(a) It may be difficult to determine whether the answers given by the child reflect the difficulty of the passage (material), or the difficulty (lack of clarity) of the questions.

(b) It may not be known how far the subjectivity or preferences (prejudices/beliefs/attitudes) of the test constructor affect the items and therefore the outcomes.

(c) It may not be known if the questions set on any passage are sufficient in number to adequately sample the content of the passage or are sufficient in scope to be an unbiased sample of all the questions that could have been asked.

(d) As construction of these tests is time consuming it is most unlikely that the average classroom teacher will/can spend the time required to write carefully constructed items and expose them to expert editorial scrutiny (as suggested by Wesman, p.111).

As a result it is possible that the difficulty, the reliability, and even the validity of such tests are likely to vary from any one teacher to another and from any one time to another. Hence, there is no certainty as to what a score of 75% or 90% on these tests might really mean - no certainty as to whether they are accurately predicting frustrational, instructional or independent levels of reading.

It is in this context of doubt that the cloze procedure (Taylor 1953) has been introduced as a viable solution to a measurement problem. This procedure involves the deletion of words from a passage of prose and the

measurement of the ability of the individual to replace these omitted words. (The procedure is reviewed in detail in Chapter 2)

Because the cloze procedure asks no questions, involves no memory component, is constructed by the simple and objective mechanical deletion of words, and ".... does not appear to be measuring a student's familiarity with the content of the passage" (Simons, 1971, p.347), it has been seen by some researchers (e.g. Bormuth, 1967, 1968; Rankin and Culhane, 1969; and Anderson and Hunt, 1972) as a realistic alternative to the problems posed by the potentially inaccurate multiple-choice testing criteria for the measurement of passage performance. With the cloze procedure, it is claimed, it is possible to have the advantage of direct testing of the individuals ability to comprehend the material associated with the use of an accurate and objective measure of this comprehension.

The purpose of this thesis is to investigate the effectiveness of the cloze procedure for this purpose. Anderson (1971a) claims it to be " ... one of the most promising techniques to emerge in recent years for measuring comprehension and reading difficulty, (p.181) whilst Klare (in Groff, 1971) believes it to be " ... clearly one of the most, if not the most, convenient and widely applicable techniques ever suggested for studying text." (p.677) These claims need to be explored in the context of the measurement of passage performance.

The next chapter will deal with the rationale and effectiveness of the cloze procedure, and will review its use by researchers as a means of predicting the suitability of reading materials.

CHAPTER II

THE CLOZE PROCEDURE AND PARAGRAPH PERFORMANCE

Although as recently as six years ago it could be written that the cloze procedure was familiar only to a small number of reading and language specialists (Spache, 1968), its use over recent years has developed greatly (see, e.g. the bibliographies of Boyce, 1973a, and Klare, Sinaiko and Stolurow, 1972). Not only is the cloze being extensively used by researchers in reading and language, but its inclusion in some reading texts (e.g. Fry, 1972, and Strang, 1968) and the numerous articles explaining its practical usefulness for classroom teachers (e.g. Anderson, 1968; Bortnick and Lopardo, 1973; Culhane, 1970; Galloway, 1973; Guice, 1969; Humphreys and Kay, 1971; Mork, 1971; Oller, 1972; Oller and Conrad, 1971; and Weintraub, 1968) have made it a potential measuring tool for the practising classroom teacher.

The procedure, which was introduced by Taylor (1953), involves the mutilation of passages of prose by the deletion of words on some mechanical basis. Introduced as a means of determining readability of material, it has been used for a wide variety of purposes over the years (see, e.g. Bickley, Ellington and Bickley, 1970, and Boyce, 1973b, p.34.)

Rationale

In introducing the cloze Taylor drew on Miller's (1951) work in communication theory, Osgood's (1952)"dispositional mechanisms" and the principles of random sampling. He chose the name 'cloze' as a derivation from the Gestaltist law of closure - the principle that behaviour or mental processes tend towards completing or 'closing' as far as circumstances permit.

The procedure as introduced by Taylor was to systematically delete words in a passage of prose and evaluate the success the reader had in accurately supplying the missing words. He reasoned that if the individual could understand the message when words were deleted, and could replace the words exactly, he was experiencing a form of closure. In order to make these cloze responses the individual had to decide from the context that remained what the missing parts were. Therefore the reader was required to have an adequate grasp of the language structure on the page as well as a grasp of the basic tone and substance of the passage. Thus Taylor claimed that the procedure provides "... 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." (1953, p.432)

Anderson (1971a) maintains that there is little empirical evidence for the explanation of 'closing' broken language patterns in the same way as one 'closes' an incomplete circle. He proposes that a more defensible rationale lies in current communication theory. The deletion of words is seen as 'noise' and the reader's task is seen as that of reconstructing the language patterns by making the most likely replacement in the light of his language system and the grammatical and semantic cues that are available. (See Figure 1).

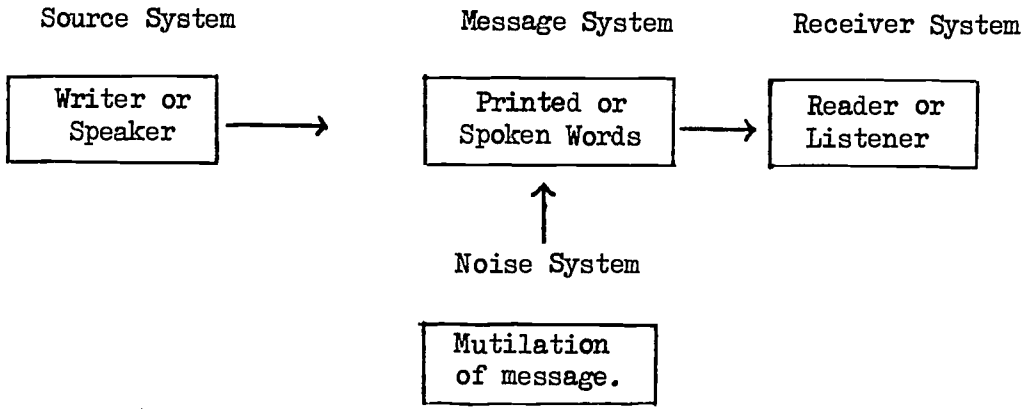

Figure 1 A Model for the Language Correspondence of a Source System to a Receiver System. (Anderson, 1971, p.179)

However, although Taylor sees the replacement of missing words as 'closing', and Anderson sees it as eliminating 'noise', these different interpretations appear to have no practical implications.

Clark and Johnson (1973) argue that on the basis of Taylor rationale "... the cloze procedure could produce spurious comprehension scores for poor readers since some words substituted will simply reflect automatic response to grammatical patterns rather than appreciation of the full meaning of the sentences or language units involved." (p.15) This will occur, they argue, because complete use of all contextual clues is not necessary for the replacement of 'functional' or 'structural' words (e.g. pronouns and prepositions), and these are easier to replace than 'content' words such as nouns, adjectives and verbs. MacGinitie (1966) also points out that missing words can often be restored correctly without "understanding" of the passage because all that is needed is a recognition of familiar patterns of expression. He feels that unless the blanks in the cloze test are appropriately selected, the cloze scores may be more a measure of language redundancy than of comprehension. This matter of the ease of replacement of various parts of speech is explored in the pilot study reported in Chapter 3 of this thesis.

It should be noted that the cloze procedure is not the same as 'fill-the-gap' or 'sentence-completion' exercises. Typically these exercises are used to gain a measure of a person's knowledge of specific and usually independent points of information, and therefore the deletions are chosen quite subjectively. On the other hand cloze procedures are mechanical and therefore objective, the concern being with a contextually related series of deletions rather than with isolated ones.

Methodological considerations.

Although consistently referred to as a simple procedure, a survey of the literature indicates that a wide variety of practices are used in the construction of tests as well as in the scoring. Taylor's (1953) introduction was a completely mechanical procedure of choosing words to be

deleted on a random or every n^{th} word basis, and calling for the exact replacement of the deleted words. However, subsequent developments have varied a number of different factors to the extent that it is difficult to talk about the cloze procedure, and which make it important that those who report research using the cloze indicate precisely what method they have used. The following sections attempt to discuss and clarify some of the variations that have occurred.

Frequency of word deletions.

There are two commonly used word deletion approaches, viz., random deletion and n^{th} word deletion, although the latter is far more common in the reported research. Amongst those who have used n^{th} word deletion the deletions usually vary between every fifth and every tenth word. Culhane (1970) suggests that every tenth word should be used with textual materials laden with fact, but that a count as low as every fifth word may be used satisfactorily with narrative materials. Probably the greater majority of researchers use an every fifth word deletion on the basis that MacGinitie (1961), in an investigation into contextual constraints in English prose paragraphs, found that the influence on word choice appears to decrease rapidly with distance of the context, and that after about five words distant the context has relatively little effect on the choice. Johnson (1968) and Anderson (1969) come to the same basic conclusion. Kerr (1970) reports that Anderson (1969) and Kerr and Smith (1968) have found that an every eighth word deletion pattern worked successfully with Australian primary school children, although he does not elaborate on the statement. The present author has found, in unpublished and unreported investigations, that younger primary school children doing cloze exercises for the first time find the deletion of every fifth word a rather daunting experience.

Most published investigations give no reason for the choice of every fifth, sixth, seventh, etc. word, and it would seem that in many cases the choice is purely an arbitrary one. In the cloze exercises devised for this investigation two different deletion patterns were used. In the Pilot Study (see Chapter 3) an every sixth word pattern was used. In the major

study an every seventh word deletion pattern was used. These deletion patterns were chosen as a compromise between the Clark and Johnson (1972) and the Anderson (1969) eighth, both of which had been found satisfactory with Australian children, and the majority of other studies which use an every fifth word deletion pattern.

As Clark and Johnson (1973) say, "... a more rigorous and selective deletion system is warranted, if proper account of contextual constraint is to be taken and children's errors in replacing deletions are to have any practical significance in relation to particular passages." (p.17). Certainly there needs to be research carried out to determine what are the most effective deletion systems according to the age of the person doing the exercise, and according to the type of content of the material.

Clark and Johnson (1973) also raise the issue of wildly fluctuating difficulty levels associated with a random deletion approach. The major purposes of the Pilot Study reported in Chapter 3 are to see if there are fluctuating difficulty levels with an n^{th} word deletion pattern, and to see if, for example, an every seventh word deletion pattern is used that there is no reason to believe a cloze exercize deleting the first word and then every seventh will necessarily be of equivalent difficulty to one starting with the deletion of the second, or the third, or the fourth, etc. and then every seventh word thereafter.

Finally in this section on frequency of word deletions there is a related question that should be considered, viz. just what constitutes a deletion element. Jongsma (1971) indicates that although researchers usually answer this on a logical basis, there is in fact no research evidence available for guidance. Thus, e.g., should numerals be subject to deletion and should hyphenated words be treated as single units or broken up into their separate parts? Klare, Sinaiko and Stolurow (1972) state that a word is usually defined "by the white spaces separating it from other words (e.g. don't, U.S.A., 2, 182, and re-enter would all be single words). Commas, apostrophes, and hyphens should be deleted along

with the rest of the word." (p.85) They believe that hyphenated words should be deleted as units only when one of their elements represents a bound rather than a free morpheme, as, for example, the co- in co-chairman. In the cloze exercises used in this study numerals were included as units for deletion, and hyphenated words were broken up into their separate parts.

Type of words deleted.

Although introduced as an n^{th} (any word) or random deletion procedure, a number of researchers have carried out investigations using specific word deletions. In his second study (1957) Taylor used three types of word deletions: 'any' words, 'hard' words (adverbs, verbs and nouns) and 'easy' words (e.g. pronouns and articles). He found that for some purposes, e.g. measuring prior knowledge of technically worded material, the deletion of 'hard' words was the best measure. However, for most purposes he found the 'any' word deletions were superior to the other forms of deletion. Greene (1965) modified the procedure by restricting words eligible for deletion to nouns, verbs, adverbs and adjectives. Louthan (1965), whilst using a purely mechanical form of deletion for part of his study, used a number of specific deletions such as proper and common nouns, as determined by morphology and syntax; specific verbs exclusive of function verbs; and specific modifier, adjective and adverb, all on a ten per cent deletion basis. He found that with all the classes listed above the lexical and grammatical redundancy was not great enough to bridge the gaps in the prose.

Rankin (1958) refers to the any word deletion by mechanical n^{th} word as structural deletion, and by specific word type as lexical deletion. He assumes that passages comprising lexical deletions measure the understanding of substantive content, while structural deletions involve an understanding of the inter-relationship of ideas and are more highly influenced by intelligence. Although Jongsma (1971) admits that there is some evidence for the psychological reality of this dichotomy, he also maintains that it is not as convincing as many would have us believe. He

believes that there is as yet insufficient evidence to suggest that the distinction applies equally well across all age and grade levels and across all types of reading materials.

Schlesinger (1968) is also critical of the structural-lexical concept. He believes that it does not take into account the deep versus surface structure of the sentence. Although the example he gives (1968, p.154) is rather extreme - he uses an every second word deletion - his point is worth considering, that instead of continuing to rely on the grammatical elements of the sentence, conventionally defined by establishing word classes or by using parts of speech and word categories, an attempt should be made to focus on the linguistic variable of word order or sentence structure.

Ohnmacht, Weaver and Kohler (1970) explored the relationship between the cloze and closure in a factorial study. They used four types of deletion systems defined as follows: "structural", "lexical", "abstract nouns" and "concrete nouns". Factor analysis identified a number of patterns which differentiated the tests. The cloze tasks could be broken into two dimensions: (a) the 'lexical' and 'abstract noun' deletions were more closely related to vocabulary and, (b) the 'structural' and 'abstract' noun deletion forms separated out in another dimension. None of the closure tasks had a major relationship with the experimental cloze tests, and the latter showed a positive relationship with performance on the associational tasks.

They then suggest that -

> 'the fact that responses to cloze tasks reflecting essentially gross deletion strategies align themselves with crude measures of comprehension does little to throw light upon the fundamental nature of comprehension other than to indicate that one can measure what passes for comprehension in more than one way...'(p.215)

and continue -

> 'Rather than standardizing a particular cloze deletion type, exploration of a wide range of deletion types which are related to particular linguistic and psychological hypotheses is needed.' (p.215)

Bowers and Nacke (1971-72) believe that the generative transformational theory of Chomsky (1957, 1965) means that one needs to re-appraise some of the use of the cloze procedure. They point out that although some researchers, e.g. Rankin (1959), Weaver (1965) and Treisman (1965), have modified the raw cloze procedure to allow for the differences between structural and referential morphemes, these attempts have not overcome the considerable problems generative theory presents for the theoretic basis of the cloze procedure. Bowers and Nacke present a tentative algorithm for the deletion of redundant words in the English language which they believe can form the basis of restitution tests "which will be both valid and illuminating" (p.31). As yet there has been no reported research using this algorithm.

Despite the doubts recently cast by linguists, the cloze procedure continues to be used, mainly on the basis of an 'any word' deletion. Although Taylor himself used specific word deletions in his 1957 study he maintained that for readability purposes to "restrict deletions to particular kinds of words is to ignore the fact that those kinds of words may not occur equally often in different materials. The difference of frequency of occurrence may itself be a readability factor; if so, its effect should be included in - not excluded from - the results." (1957, p.25) On the other hand Clark and Johnson (1973) suggest that this might be just as much an argument for carefully analysing the passages before determining the type of deletion.

In the investigations carried out by the author for this thesis only mechanical 'any word' n^{th} deletions were used - the cloze procedure as originated by Taylor (1953). However, the purpose of the Pilot Study reported in Chapter 3 was to investigate the extent to which different types of words may not occur equally often in different cloze versions of the same material.

Scoring Cloze Tests

It is usual for cloze tests to be scored for exact replacement of the deleted words, although various other methods have been explored.

For example, Guice (1969) graded on the basis of two points for exact replacement and one point for a synonym. Weintraub (1968), although reporting that most research has been carried out using exact work replacement, suggests that synonym replacement is allowable. Miller and Coleman (1967) using two scoring methods, viz. (a) exact replacement, and (b) 3 points for exact replacement, 2 for a synonym and 1 for correct part of speech, found a correlation between these methods of 0.99. This, together with the evidence of research, e.g. Taylor (1953), Rankin (1957), Ruddell (1964) and Bormuth (1964, 1965), suggests that not only is scoring for exact replacement simpler and more reliable (as no subjective assessments have to be made as to what are allowable alternative replacements), but that scoring allowing for synonym replacement does not lead to better discrimination between individuals.

On the other hand it could be reasonably argued that if the cloze test was being used to consider the individual's performance rather than to assess his performance relative to others, that some purpose might be achieved by scoring for synonyms and logical replacements. Schoelles (1971) believes that when the procedure is being used for measuring student ability the scoring of synonyms is desirable. She argues, for example, that enriched vocabulary use - such as 'constructed' for 'made' - should not be penalized. Boyce (1972) in a study in which responses to a cloze test were scored for (a) exact replacement, and (b) exact replacement or synonym or logical replacement, found that the mean scores increased from 21.29 for scoring method (a) to 28.78 for scoring method (b). An investigation of the accepted synonyms and logical replacements indicated that, in some cases at least, the exact replacement word was not the common usage word of the children.

Oller (1972) argues that although mean scores tend to be higher when acceptable substitutes are allowed, the increase in total test variance is so small as to be scarcely worth the extra effort involved, and Bormuth (1965) suggests that exact word replacement is required for validity.

For practical purposes it is probably reasonable to maintain only the exact replacement scoring system as any other method loses objectivity, and the work involved in determining what are acceptable synonyms and logical replacements is considerable. One possible solution is the development of clozentropy (Darnell, 1970). Clozentropy was developed as a procedure for testing English language proficiency of foreign students. It has amongst its theoretical assumptions one that states "... that a measure of proficiency in language should index one's ability to conform to existing group norms of language rather than to some prescriptive model or idealized language pattern." (p.36) Thus, although Darnell uses the cloze technique, he also uses an entropy measure which indexes the compatibility of an individual's responses with those of a selected criterion group. This leads to a scoring system that is mathematically precise, which avoids entirely the right/wrong judgements on an item by item basis, but which is rather complex.

In the two studies carried out for the purpose of this thesis only exact replacement scoring was used, mainly because in both cases the results were being compared with, or related to, other studies using exact replacement scoring.

Number of deletions.

Kerr (1970) points out that because random or n^{th} word deletions lead to a number of non-discriminating items being included, the reliability of the test is lowered if there are only a few items. Thus the test has to be long enough to be reliable, but not long enough to cause fatigue and boredom. Taylor (1956) suggested that 50 items led to a stable score, and this was supported in principle by Bormuth (1964). In a later study (1965b) Bormuth presented a table, based on Lord's (1955) formula for standard errors, which allows an estimate of the standard error to be made according to the number of deletions and the number of subjects.

Obviously the length of the test is affected by the rate of deletion.

Thus a fifty item every fifth word deletion exercise would be much shorter than a fifty item test with every eighth word deleted. Although Anderson (1971b) suggests that it makes very little difference "except in terms of efficiency and reliability", (p.38) whether one uses every sixth, seventh, or eighth word, it does, of course affect the total length of the passage being used. There appears to be no consistent body of research to indicate what is the best deletion pattern to use according to the age of the child and the type of material, and hence there is a possibility that longer passages may be needed for younger children than with older ones. Related to this is the matter of motivation. There appears to be no research on how performance on the cloze affects the motivation of the child to continue. It could be hypothesized that the smaller the number of words between gaps (deletions), the more difficult it is for the younger child.

In all the tests devised for use in this thesis a fifty item cloze was used.

Related to the question of the length of tests, and a question rarely mentioned in discussions of the procedure, is that of whether a 'run-in' should be used before the deletions actually commence. There appears to be some confusion on this matter. Some researchers start deletions from the first sentence, others leave the first sentence or two, whilst others leave as much as the first paragraph of the material before commencing deletions. Oller (1972) writes, "As is customary, the first and last sentences of each paragraph were left intact." (p.152). Klare, Sinaiko and Stolurow (1972) state that although some writers suggest that no words be deleted from the first and last sentences of a passage, they feel it to be unnecessary "... except for subjects like young children, near-illiterate adults, or such who need a great deal of help." (p.85) Anderson, (1972 - personal communication) writes, "Your suggestion of leaving an initial paragraph unmutilated is sound though depending on the length of the paragraph it may not be necessary to leave the whole paragraph ... but it is important, I agree, to set the scene so-to-speak."

Another factor not usually mentioned, but of practical importance, is whether the subjects read through the material first before attempting to replace the deleted words. This, together with the 'run-in' factor, may have some influence on the strategy the subject uses, and therefore on his score. It is feasible that leaving the initial paragraph, or at least part of it, unmutilated, together with the instruction to read through the whole task first before attempting to fill in the gaps, would allow the subject to approach the task as a whole because he has a better grasp of the total context, mood, style, etc. On the other hand, if he simply starts at the beginning without any overview, replacing each omission as he comes to it, he may treat the passage as a series of sub-tasks, as a series of bits of information. If this does in fact happen it could account for some of the replacements which, although patently wrong in the total context of the passage, make sense in the context of the few words immediately preceding and immediately following the particular deletion.

Anderson (1972 - personal communication) states that the usual instruction is to read through the whole passage and then fill in the missing words. Klare, Sinaiko and Stolurow (1972) and Bormuth (1964b) do not mention this in their sets of instructions, nor indeed does Anderson (1971a). However, in another article, Anderson (1971b) gives the following instructions for the use of the cloze with primary school children:
" ... I want you to read each story and guess the missing words. Then I want you to print in each space the <u>one word</u> you think should go there." (p.39)

The majority of journal articles do not mention the instructions given. This would seem to be an unfortunate omission as it may have a major influence on the strategy/strategies used by the children to do the exercises. There, in fact, seems to have been very little research carried out into the strategies used by the subjects. Jenkinson (1957) selected high school students who had done very well, or poorly, on cloze tests and asked them to verbalize their reasons for the insertion of words on another test. These verbalizations were then analysed and showed that the

higher scoring students were much better in recognizing syntactical cues, sensitivity to style, language structure etc. Although there would be problems involved, it would seem feasible to ask children to introspect about the way they went about the task, to identify the strategies used, and then to test them experimentally by using varying forms of instructions.

The instructions used in the testing carried out by the author for this thesis can be found on page 53. These instructions include the sample exercise given to acquaint the children with the procedure. All the cloze exercises devised included a 'run-in' before deletions commenced, the length of which varied from exercise to exercise. The actual length of the 'run-in' was determined to some extent by the fact that in the major investigation the exercises were photostats of the original text from the book and as far as possible the exercise was kept to one page only.

Format of the exercise.

General practice is for exercises to be compiled by typing out the passage and replacing the deleted words with a blank space of standard length, usually ten or fifteen typewriter spaces. The tests are then presented in duplicated form with the subjects writing the replacement words in the spaces provided. Unless one has a see-through template with the correct replacements written on it, the correction of cloze exercises compiled in this manner can be very frustrating. An alternative is to number the spaces and have numbered blanks of standard length on the right hand margin of the page, or on a separate answer sheet. This method greatly facilitates correction as a simple vertical answer card can be used to match up answers. There may be one possible disadvantage however in that subjects have to search for the correct place to commence each time after having written the answer in another place and may thus lose the thread of the passage.

Anderson (1971a) has suggested another format. He claims that his research has shown that blanks of the same length as the deleted word are

an effective alternative. He therefore suggests that cloze exercises can effectively be constructed by glueing paper over the words in the original that are to be deleted, and then photocopying the passage. Such a method would mean that the size of print, length of deleted word, illustrations, and page layout could be contextual cues involved in the exercise. Although Klare, Sinaiko and Stolurow (1972) claim that standard size blanks should be used and that use of blanks of the same size as the deleted words provides undesirable cues, it would seem reasonable to use whatever cues the materials can give. After all, what we are trying to determine is whether the child (subject) can comprehend the material - as it is in the book.

The present author used a variation of Anderson's photostat format for the main investigation in this thesis. In all there were 112 different cloze exercises. To have produced these in typewritten duplicated form would have been very costly. The cutting out or pasting over of words to be deleted turned out to be a very frustrating and time consuming task. Instead, words to be deleted were obliterated by the use of white liquid retype. Although Anderson suggests that the students can write in the answers in the spaces left in this photostat format, this author found that the space left with many of the small type forms together with the general size of primary children's writing, made this impractical. Thus each of the whited-out blanks was numbered and a separate answer sheet provided. (See Appendix H) For the pilot study, which was based on the work of Clark and Johnson (1972), the same format as they had used was used, viz. the passage was duplicated, with blank spaces of constant length, numbered, and numbered blanks were provided on the right hand margin of the page. (See Appendix D)

Close procedure and paragraph performance: A review of research.

The major problem facing the use of the cloze as a means of replacing multiple-choice tests as a measure of paragraph performance has been the lack of a frame of reference by which scores on a cloze test might be

interpreted. Although a higher score for one individual obviously indicates that he has performed better than one who has obtained a lower score, the absolute figures (i.e. the raw cloze percentage scores) do not tell us how well the readers comprehend the material. Likewise, it is reasonable, on the surface at least, to say that a higher mean score for one set of material indicates that it is of an easier standard than material that obtains a lower mean score, but this does not tell us much about the actual difficulty of the material.

In order to overcome this problem attempts have been made to determine comparable cloze and multiple-choice comprehension test scores, especially in relation to 75% and 90% levels of comprehension. By doing this it is believed that passage performance criteria can be established that will allow teachers to use the simpler, mechanical and objective cloze procedure, rather than the subjective, problem-ridden multiple-choice process.

Bormuth (1967)

The earliest work in this area was carried out by Bormuth (1967). In this study a 50 item cloze test and a 31 item multiple-choice test were made over nine passages. Each of the multiple-choice tests contained questions thought to measure seven different types of comprehension skills. Validation was tested by asking two qualified test experts to independently classify the items as to type and to discard items, and also by trying out the items on 73 children and discarding those items that were negatively correlated with the total.

The passages each contained approximately 275 words and had a Dale-Chall readability from 4.5 to 6.5. The exercises were administered under untimed conditions to 100 pupils in grades 4 and 5. In each case the cloze form of the test was administered first, the multiple-choice form being taken three days later.

Scores for each individual over all nine of the cloze and multiple-choice tests were summed to form two sets of scores. A scatter plot of the

two sets indicated linearity. The product moment correlation was then calculated and the data fed into regression equation to calculate the most probable multiple-choice score associated with each of several cloze scores.

The results indicate that if the conventional passage performance criteria are accepted, a passage on which a student receives a cloze score of 38% is sufficiently understandable to him to be used in his instruction - i.e. a score of 38% on a cloze test is equivalent to a score of 75% on a multiple-choice test over the same material. Likewise, a 50% result on the cloze is equivalent to 90% on a multiple-choice test. Bormuth also provided comparable scores if one demands as a criterion a multiple-choice equivalent score corrected for guessing - 43% and 52%.

Bormuth quite correctly warned that the accuracy of his predictions is only as good as the cloze test data he had collected, and that it should be clearly understood that the comparable scores hold good only where the dependent scores are obtained using test instructions and tests similar to those used in his study - although he doesn't really detail them, particularly the instructions.

Bormuth (1968)

In a follow-up study, Bormuth (1968) set out to determine a set of criterion scores comparable to scores on oral reading tests. In this study the materials used were paragraphs from the four forms of the Gray Oral Reading Tests (1963). Each form contains 13 paragraphs in a graded sequence ranging from a very easy pre-primer level of difficulty through paragraphs difficult enough to challenge able high school students. For the comprehension tests it was necessary to augment and revise some of the items in the published versions of the tests in order to obtain a reliable measure of how well students comprehended each paragraph. The items were constructed by using transformations (after Chomsky, 1957) on the language in the passages.

Two versions of a cloze test were made from each passage by deleting different patterns. Subjects were drawn randomly from grades 4-6 in a single school. Two of the four paragraphs at each level were randomly assigned to each subject who took these as cloze tests. The complementary pair were taken by each subject as oral reading tests.

Since oral reading test scores were often available for only a portion of the range of paragraph difficulty, ordinary regression techniques could not be used to determine the comparable scores. Instead a simple matching procedure was used. To find the cloze score comparable to the 75% comprehension criterion, the most difficult paragraph level on which a subject obtained a comprehension level of 75% was found, and the subjects cloze score on that level was noted. When no comprehension score of exactly 75% was obtained, the level of paragraph difficulty having the score nearest to 75% was used. The cloze scores were then averaged across subjects to obtain the comparable score.

In fact, the matching procedure used in this study was probably more defensible than the regression method in the first study, when a 'goodness-of-fit' approach would seem to have been more appropriate.

Cloze scores of 44% and 57% were found to be comparable to the criterion reference scores of 75% and 90% respectively. These can be compared with the 38% and 50% of the previous study. The seven point difference between the independent level cloze scores in the two studies can be explained - according to Bormuth - by the fact that a ceiling effect was observed in the multiple-choice scores in the earlier study, and this probably suppressed the multiple-choice scores at the upper end of the range, thus resulting in an artificially low comparable cloze score. On the other hand, the difference might be explained, at least partly, by the difference in methods of obtaining equivalence.

Whilst pointing out that the study needed replication, and that results could only be generalized to subjects and passages similar to those in the

study, Bormuth believed that any replication would obtain similar results because most of the items written were written as transformations, thus precluding the possibility of them being manipulated arbitrarily to alter their difficulties, and because most of the paragraphs were very short and the number of items written for every passage was relatively large, nearly every item that could have been written for each paragraph was used, thus reducing the possibility of bias.

Rankin and Culhane (1969)

Rankin and Culhane (1969) carried out what was essentially a replication of Bormuth's 1967 study. Although there were slight differences between the procedures used, the investigation was probably comparable in all significant aspects except that Rankin and Culhane used only fifth grade children as subjects.

Although there was fairly close agreement between Rankin and Culhane's scores and Bormuth's scores at the 75% and 90% level, there are considerable differences at other levels. (See Table 1)

In fact, taken over the range of 50% to 100% multiple-choice scores the cloze comparable scores show a range of 39 (19-57) in Bormuth's study and 65 (10-74) in Rankin and Culhane's.

TABLE 1

Equivalent cloze and multiple-choice percentage scores for Bormuth (1967, 1968) and Rankin and Culhane (1969)

Multiple choice scores	Bormuth 1967	Bormuth* 1968	Rankin and Culhane	Difference
50	19		10	+ 9
55	23		15	+ 8
60	27		22	+ 5
65	31		28	+ 3
70	35		35	0
75	38	44	41	± 3
80	42		48	- 6
85	46		54	- 8
90	50	57	61	- 4) -11)
95	53		67	-14
100	57		74	-17

* Note that in Bormuth's 1968 study comparable scores were only given for the 75% and 90% levels.

Rankin and Culhane point out that the average difference is in fact only 3.1 percentage points, but this is not very convincing. Rankin and Culhane argue that the greatest discrepancies lie in the scores comparable to multiple choice scores of 85 and above, and that this may be accounted for by Bormuth's belief that ceiling effects gave him artificially low comparable cloze scores in the upper levels. However the difference column (Table 1) with its increasing differences at the extremes of the range exhibits all the manifestations of the typical regression effect, and this is a more likely explanation of the differences.

There is reasonable correspondence between Bormuth's 1968 scores and Rankin and Culhane's scores at the 75% and 90% levels - 44/41 and 57/61. On this basis Rankin and Culhane saw fit to say:

> 'It is now possible for teachers to interpret cloze test results with some degree of confidence by using specific percentage scores as criteria of acceptable performance. The use of the comparable cloze and multiple-choice scores found in this study should be particularly useful for a teacher who wishes to measure reading comprehension of pupils in a specific subject matter field by using a cloze test based on material in that field.' (p.198)

Anderson and Hunt (1972)

The only other published study in the development of comparable cloze and multiple-choice scores is that of Anderson and Hunt (1972). This study was carried out with children in schools in Papua New Guinea who had learned English as a second language. Although Bormuth's basic approach was used there were some differences. There is no indication of the length of the passages, except that they were 'short', and whereas Bormuth (1967) and Rankin and Culhane (1969) both used 31 item multiple-choice tests per passage, Anderson and Hunt used 90 items over the nine passages used - i.e. an average of ten items per passage. There is no indication of the validation of these multiple-choice items. Also, whereas Bormuth (1967, 1968) and Rankin and Culhane (1969) had used a deletion rate of every fifth word, Anderson and Hunt used an every eighth word deletion rate.

Anderson and Hunt achieved comparable cloze scores of 44% (for 75% m-c) and 53% (for 90%), and come to the conclusion that the agreement between their scores and those carried out in a different country and within a different educational system seems remarkably close. They conclude by claiming that although the criteria they derive and those previously derived by Bormuth and Rankin and Culhane will not be applicable in all future cloze and multiple-choice comprehension tests, the results should enable primary school teachers to use their results from cloze tests with confidence to judge the suitability of reading materials for particular pupils.

TABLE 2

Summary of cloze comparable criteria for the four studies

	Multiple-choice criteria	
	75%	90%
Bormuth (1967)	38%	50%
Bormuth (1968)	44%	57%
Rankin and Culhane (1969)	41%	61%
Anderson and Hunt (1972)	44%	53%

Table 2 summarizes the comparable cloze and multiple-choice criterion scores for the four studies discussed.

Mosberg, Potter and Cornell (1968)

Relevant to the above studies is the investigation carried out by Mosberg, Potter and Cornell (1968) into the relationship between cloze and multiple-choice tests. Working at two grade levels - grades 5 and 8 - with reading passages at difficulty levels either two years below, two years above, or at subject's grade level, they tested at each grade level and each passage difficulty level a large number of reading passages with a large subject sample.

Table 3 shows the obtained correlation co-efficients between cloze and multiple-choice performance.

TABLE 3

Correlation between cloze and multiple-choice performance according to grade and difficulty level. (after Mosberg, Potter and Cornell)

Difficulty level	Grade 5	Grade 8
Low	.649	.190
Medium	.429	.367
High	.434	.247
Overall	.535	.335

The correlations reported in Table 3 above suggest that although the cloze procedure does measure some component of comprehension as measured by multiple-choice tests, there is a large component measured by the multiple-choice tests which is not accounted for by the cloze. Mosberg et al do point out that their correlations were calculated on the basis of matched pairs, and that insofar as these were not perfect the correlations are depressed. However they do feel constrained to say that they are cautious in their acceptance of the cloze procedure as a predictor of what a student would score on a multiple-choice test. This study deserves replication.

Bormuth (1971)

The four studies reported above are based on acceptance of the frustration, instructional and independent levels of reading. In fact there appears to be no empirical evidence to support these three levels, i.e., that although the traditional criteria of 75% and 90% have been widely accepted by reading researchers and teachers, there is no evidence that they are any more than operationally defined levels. Powell (1968), Hunt (1969), and Spache (1969) consider these Killgallon-Betts Criteria to be arbitrarily fashioned and not commensurate with reality, although their major argument is with word recognition criteria rather than with the multiple-choice comprehension criteria discussed in this paper. Spache however believes the 75% for instructional reading level should only be about 60%.

Because he believed these criteria, if not arbitrary, were at least unexplicit and unrationalized, Bormuth (1971) set out to establish rational passage performance criteria using the cloze procedure.

Bormuth believed that a reasoned approach to identifying the criterion level of performance on a passage would set the score at the performance level where a weighted sum of the outcomes showed that a maximum benefit was to be expected. In studying the variables affected Bormuth suggested that the following were relevant: Cognitive variables such as learning and retention and transfer of information in the passage; Proficiency

variables such as rate of reading and latency of responses acquired from the passage; Affective variables such as students' preferences for the subject matter, style, the difficulty of the passage and the students willingness to study it; Economic factors such as the costs involved in preparing suitable materials; and Psychosocial factors such as the effects on self concept of having to study materials at the given level of difficulty relative to the subject's level of ability.

In the series of studies reported in his 1971 paper, Bormuth included only the following factors in his criterion selection model: Measures of information gain, rate of reading, willingness to study, preferences for the subject matter, style and level of difficulty.

In these studies he set out to establish:

(a) the regressions between each of these variables and cloze scores;
(b) a set of weights representing the relative values placed upon each of these variables;
(c) what variables influenced the shapes of the regressions and therefore required a differentiation of the passage performance criterion score.

Initially the studies were designed to permit the results to be generalized to students in grades 3 - 12, to materials on most of the topics and at most of the difficulty levels that these students would be likely to encounter in instruction, and to each of the major purposes for which students are likely to read a passage. However, because cloze and grade level consistently interacted in all the regressions, it was necessary to identify different criterion scores at each grade level. Also because students assigned different ratings to materials depending upon whether they were to be used for textbook, reference, or voluntary reading purposes, it was necessary to allocate criterion scores for each of these three purposes at each grade level.

As a result, Bormuth comes up with a set of scores for each grade level as shown in Table 4. Only grade 3 scores are used here for illustrative purposes.

TABLE 4

Cloze scores and dependent behaviour efficiency for three reading purposes at grade 3 level (Bormuth, 1971)

Criterion	Cloze score	Dependent Behaviours				
		Info gain	Rate rdg	Subject matter	Style	Difficulty
Textbook	54	81	59	100	99	55
Reference	52	78	57	100	98	47
Voluntary	62	90	68	97	99	1

The figures in Table 4 are interpreted in this way: a cloze score of 54 on a passage from a textbook may be regarded for grade 3 children as producing an efficiency rate of 81% on information gain, 59% on rate of reading, 100% on subject matter, etc. Bormuth does not make very clear what he means by efficiency rate, and although an 81% efficiency rate on information gain seems a reasonable statement, 100% efficiency rate for subject matter or 99% efficiency rate for style, is not readily meaningful.

Bormuth believes that although the scores he presents are only a crude first approximation to those ultimately sought as passage criterion, they are probably much superior to any other passage criteria in use. Thus, whilst cautioning practitioners and researchers about using them without considerable caution, since they contain both systematic and random error, he does suggest that they be used.

Summary

Descriptions have been given of two different ways in which the cloze procedure has been used to obtain passage performance criteria:

(a) By establishing comparable cloze scores for multiple-choice test performance (Bormuth, 1967, 1968; Rankin and Culhane, 1969; and Anderson and Hunt, 1972). The assumption lying behind these studies is

that if you accept the traditional 75% and 90% levels of performance as indicating instructional and independent levels of reading, it is better to use the established equivalent cloze scores as the measure, because the cloze method of measuring comprehension is simpler, the mechanical deletion of words is an objective procedure and does away with all the problems associated with subjectivity and the difficulty of items in multiple-choice tests.

(b) By establishing completely new passage criteria and using cloze scores as the direct measure. (Bormuth, 1971)

Research implications.

Whether one accepts the approach of (a) or of (b) above, in both cases the criterion score is established as a single score. For example, if one takes Bormuth's 1968 criterion of 44% as indicating the instructional level, this means that if a child scores less than 44% on a passage thought to be representative of that material, then the passage is too difficult for him, or if he scores 44% or above, it is of suitable difficulty.

There are two problems associated with this approach:

(a) The assumption is that any one cloze test constructed over a given passage of material is equivalent in difficulty to any other cloze test constructed over the same passage. If the cloze deletion pattern used is an every fifth word deletion, there are five possible cloze tests that can be constructed, if every seventh word, there are seven possible cloze tests, and so on. As there is no necessary consistency in the English language as to the length of sentences, the position of words in sentences, and the relationship of words to one another within sentences, it does not necessarily follow that any one deletion pattern will be of the same difficulty as any other deletion pattern within the same paragraph. This would not matter if the scores were being used simply to rank the children doing the test in some particular order, but when the score is being used to relate the performance of the child to a single score criterion, then the actual difficulty of that particular cloze test as compared to any other

cloze test that could have been constructed over the same material is a question that needs to be answered.

The purpose of the pilot study (Chapter 3) is to investigate this matter.

(b) Secondly, the use of a single score criterion for any material and any deletion pattern, suggests a precision that is unreal. The purpose of the main investigation (Chapter 4) is to establish operationally the range of appropriate scores rather than a single criterion score.

CHAPTER III

PILOT STUDY

The purpose of the pilot study reported in this chapter was to determine if it could be predicted that any one of the possible alternative cloze forms of a passage could be significantly easier, or more difficult, than the other forms.

Factors determining difficulty of word replacement.

There are a number of ways in which deleted words could be categorized in terms of their possible difficulty of replacement.

Parts of Speech.

Parts of speech influence comprehension (Huus, 1968), and Bormuth (1966) has shown that the ratio of pronouns to conjunctions is a good predictor of difficulty. Louthan (1965) found that if prepositions, conjunctions or pronoun substantives are deleted, there is no appreciable difference between the performances on tests following the cloze materials and those following unmutilated passages, whereas specific verb deletions, noun and modifier deletions lead to marked loss in comprehension. Elley (1969) found, with a sample of secondary school students, that prepositions and pronouns were the easiest to replace in the particular cloze exercise he used, and that nouns were the most difficult. In fact, as a result of his studies, he proposed a noun frequency count as an appropriate means of determining the 'difficulty' - and hence the readability - of reading materials.

Length of Word.

Long words have often been thought to be more difficult, and the number of syllables is a common element in many of the generally accepted

readability formulas (See p.6). Coleman (1967) found high correlations between difficulty and number of letters, number of syllables, and number of affixes, stems and inflexional morphemes. Correlations found between number of syllables and passage difficulty include 0.44 (Gray and Leary, 1935), 0.69 (Flesch, 1950) and 0.63 (Bormuth, 1966).

Familiarity of Words.

If it is assumed that meaningfulness is largely an outcome of frequency of exposure then it can be argued that the comprehension difficulty of a passage will be strongly affected by the number of unfamiliar words included. Some support for this is given by a number of studies. Dale and Chall (1948) in their reading difficulty study found that of the five indices they used the highest correlation with their criterion was the proportion of words outside the Dale list. Spache (in Hunnicut and Iverson, 1968) obtained a correlation of 0.68 in a similar study. Gray and Leary (1935) found that the factor most closely correlated with reading comprehension for poorer readers was the number of familiar words in the material. Lorge (1948), Forbes (1952) and Bormuth (1966) have all found similar relationships between familiarity of words and difficulty of comprehension.

Elley (1969) suggests that this relationship is further strengthened by the fact that the measure of familiarity is relatively weak. The words in the passages are classified as either familiar or unfamiliar, with no intermediate categories. "Since correlations which depend on a two-unit scale are usually lower than those based on a graduated scale, it would seem logical to conclude that a more refined measure of familiarity would make for an improved predictor of readability." (p.414)

Word categories used in this study.

The evidence discussed above suggests that there is justification for investigating the ease or difficulty of replacing deleted words. For this purpose four word categories, each with sub-divisions, were determined.

(a) The number of letters per word. This was sub-divided into four; 1-2, 3-4, 5-6, and 7 or more letters.

(b) Number of syllables per word. This category was sub-divided into three; 1, 2, 3 or more syllables.

(c) Whether the word was 'In' or 'Not in' common word lists. For this purpose a composite list of 'key', 'basic', 'instant' and 'sight' words compiled from the lists of Edwards and Gibbon (1964), Fry (1968), Kucera and Francis (1967), McNally and Murray (1962) and Rinsland (1945) was used. In total this list included 344 words, including words such as 'and', 'but' and 'came', which were in all five lists, and words such as 'woman', 'those' and 'yet' which were included in only one of the lists. The complete composite list is included as Appendix B.

(d) The part of speech. Eight sub-divisions were used; adjectives, adverbs, articles, conjunctions, nouns, prepositions, pronouns and verbs.

Determination of 'easy' categories.

Clark and Johnson (1972, Appendix B, pp. 23-25) report in detail part of the results of their investigation. Included are the percentage errors made by a sample of 55 grade 6 children in Victorian metropolitan schools in replacing words deleted from a passage from Doug of Australia (Cavanna, 1965). For this cloze exercise every eighth word, commencing with the first, was deleted. This was the data used in this particular pilot study to determine the 'easy' categories of words to replace, and then to predict what would probably be the 'easiest' and 'most difficult' of the possible alternative cloze forms of the passage. The Clark and Johnson data is shown in Appendix C.

Table 5 shows the percentage of correct responses for each of the sub-divisions of each of the four chosen categories using the data of Clark and Johnson (1972). The figures indicate quite clearly that for these subjects, with this particular passage, the easiest sub-categories of words to replace were those words that were 1 or 2 letters long, and/or were of one syllable,

and/or were 'In' common words lists, and/or were 'structural'/'functional' words - articles, conjunctions, prepositions or pronouns.

TABLE 5

Percentage of correct responses in each of the four categories.

Category	Percentage
1. Length of words	
1-2 letters	61.9
3-4 letters	47.1
5-6 letters	36.5
7 or more letters	21.1
2. Number of syllables	
1 syllable	48.8
2 syllables	26.3
More than 2 syllables	24.1
3. Words in common words lists	
In lists	52.2
Not in lists	27.1
4. Parts of Speech	
Adjectives	24.8
Nouns	38.2
Adverbs	38.8
Verbs	43.9
Articles	44.0
Conjunctions	51.8
Prepositions	54.5
Pronouns	67.6

Predicting 'easiest' and 'most difficult' cloze versions.

An every seventh word deletion cloze test was then prepared using the same passage. Each of the deleted words from the seven possible cloze tests

was then placed into the appropriate category. Table 6 shows the number of words in each of the subdivisions of each of the categories for each of the seven deletion patterns.

TABLE 6

Number of words in each category for each of the possible cloze versions.

Category	Pattern	Number of words 1	2	3	4	5	6	7
1. Length of word								
1-2 letters		8	10	8	15	8	10	12
3-4 letters		23	22	23	18	21	26	23
5-6 letters		5	13	10	12	9	11	10
7 or more letters		14	5	9	5	12	3	5
2. Number of syllables								
1 syllable		34	39	33	39	35	41	40
2 syllables		7	9	13	8	8	8	8
More than 2 syllables		9	2	4	3	7	1	2
3. Words in common word lists								
In lists		30	30	29	37	32	37	36
Not in lists		20	20	31	13	18	13	14
4. Part of speech								
Adjectives		5	6	5	6	7	5	5
Nouns		16	13	15	8	12	11	9
Adverbs		10	11	12	7	18	8	7
Verbs		3	2	5	6	1	6	6
Articles		4	4	3	4	4	5	6
Conjunctions		1	2	3	5	2	3	1
Prepositions		7	6	5	10	3	4	10
Pronouns		4	4	2	4	3	7	6

The information in Table 6 indicates that, if the number of 1-2 letter words, the number of 1 syllable words, the number of words in common word lists, and the number of articles, conjunctions, prepositions and pronouns

are taken as the criteria for degree of difficulty of replacement, pattern 4 should be the easiest version and pattern 5 should be the most difficult version. Table 7 compares these two patterns according to the number of words deleted which belong to all four 'easy' categories (i.e. the number of words that are 1-2 letters and are 1 syllable and are in common word lists and are either articles, conjunctions, prepositions or pronouns), those that are in any three of these categories, and so on.

TABLE 7

Number of words in the 'easy' categories for pattern 4 (easy version) and pattern 5 (hard version) from Doug of Australia.

Combination	Number of words in Pattern 4	Pattern 5
All four 'easy' categories	11	3
Any three	4	6
Any two	21	18
Any one	7	11
None of the 'easy' categories	7	12

In order to further test the ability to predict the difficulty of deletion patterns within the same passage, a passage was chosen at random from Deserts (Goetz, 1956). The excerpt has a Fry (1968) readability rating of Grade Six, and was from a primary science reference, whereas Doug of Australia was from a children's novel.

Using the same method as described above, it was predicted that deletion pattern six (i.e. deleting every seventh word commencing with the sixth word in the passage) would be easier for the children to do than deletion pattern one. Table 8 summarises the difference between the two patterns by showing the number of words in each of the 'easy' categories.

TABLE 8

Number of words in each 'easy' category for pattern 1 (hard version) and pattern 6 (easy version) from Deserts.

Category	Number of words	
	Pattern 1	Pattern 6
1-4 letters	26	35
1 syllable	36	39
In common word lists	26	33
Conjunction/article Preposition/pronoun	15	20

As a result of identifying what appeared to be the easiest and the hardest versions for each of two passages, it was decided to experimentally test the following hypotheses.

Experimental Design

Hypotheses

1. That the mean score for grade 6 children doing a cloze test on the passage from Doug of Australia with every seventh word deleted commencing with the fourth word will be significantly higher than for those doing a cloze test for the same passage with every seventh word deleted commencing with the fifth word.

2. That the mean score for grade 6 children doing a cloze test on the passage from Deserts with every seventh word deleted commencing with the sixth will be significantly higher than those doing a cloze test for the same passage with every seventh word deleted commencing with the first word.

Procedure

Subjects for the experiment were 196 grade six children from eight Melbourne metropolitan primary schools. All schools used were in south-eastern suburbs. Sex distribution was approximately equal. The Doug of

Australia passage was done by 106 children (53 for each pattern) and the Deserts passage by 90 children (45 for each passage). The difference in numbers is due to only one passage being used in one of the grades, where half the grade did the cloze test and the other half did an alternative task.

The cloze tests were prepared by the duplication method, with the blanks numbered and numbered blanks provided on the right hand side of the page. (See Appendix D for sample)

The four experimental cloze tests were randomly distributed to the children in each grade. When this had been done children were shifted so that no child was sitting next to another who was doing the same test, or a test from the same passage. This procedure was carried out because the author has found that children become aware of the fact that the answers to their deletions are in the text of the alternate form being done by the person next to them and there is therefore a tendency for some children to cheat. Thus, unless this provision is made, spurious and quite misleading results can be obtained.

The administration of the tests for this pilot study was carried out by nine third year Diploma of Teaching (Primary) students from State College of Victoria, Toorak. These students attended a briefing session with the author before going to the classrooms to administer the tests, and were also given written instructions as to the procedure to follow. (See Appendix E)

Results

Table 9 shows the mean replacement scores for each of the two experimental patterns for the two passages. As 50 words were deleted for each of the cloze tests, the highest possible score for any subject was 50.

TABLE 9

Mean number of correct replacements for the two passages.

	Doug of Australia		Deserts	
	Pattern 4 (easy)	Pattern 5 (hard)	Pattern 6 (easy)	Pattern 1 (hard)
Mean	25.71	16.39	18.10	13.70
S.D.	8.49	7.11	10.36	8.95
Variance	72.08	50.55	107.33	80.10
n	53	53	46	46

For the passage from Doug of Australia the difference between the mean scores of 25.71 ('easy') and 16.39 ('hard') was significant at the .001 level with a t of 6.128. The difference between the subjects' performances on these two cloze versions of the same passage is reflected by the fact that for the 'easy' passage 22 of the subjects obtained scores of 30 and above, whilst only two of the subjects doing the 'hard' passage obtained similar scores. The actual ranges of scores obtained were 7 - 47 for the 'easy' passage, and 1 - 34 for the 'hard' passage.

Although the passage from Deserts had been rated at Grade 6 level by the Fry (1968) readability graph, the children in the sample used for this study found it very difficult. For the predicted 'easy' pattern the mean replacement was only 18.1 (a replacement rate of only 36.20%), with a relatively large standard deviation of 10.36 and a range of scores from 7-47. For the 'hard' pattern the replacement rate was only 27.4%, and if the two highest scores for this pattern, which were 13 above the next highest score, are taken out, the mean correct replacement rate falls to 24.6%.

However, despite the overall difficulty of this passage, the hypothesis that the 'easy' pattern would yield a significantly higher mean replacement score than the 'hard' passage was supported, with a t of 2.180 which was significant at the .05 level.

For both the passages chosen for this pilot study, the mean score obtained by the children doing the predicted 'easy' pattern was significantly higher than that for the children doing the predicted 'hard' pattern. Thus both of the hypotheses are supported.

For the detailed investigation of the rate of correct replacement of words according to the four categories, only the results of the Doug of Australia passage were used. The Deserts passage was not used because the overall difficulty was such as to suggest that insufficient useful information would be obtained to warrant the time involved in a detailed investigation of the results.

Table 10 shows the percentage of correct replacements for each of the categories for the two experimental cloze tests from the Doug of Australia passage. These figures clearly indicate support for the predictions made regarding the difficulty of replacing deleted words.

TABLE 10

Percentage of correct replacements for each category for students doing the two experimental patterns from Doug of Australia.

Category	Percentage correct replacements	
	Pattern 4 (easy)	Pattern 5 (hard)
1. Length of words		
1-2 letters	73.2	56.5
3-4 letters	58.6	36.8
5-6 letters	40.6	26.7
7 or more letters	25.5	16.6
2. Number of syllables		
1 syllable	61.2	39.1
2 syllables	31.9	15.3
More than 2 syllables	36.2	22.4
3. Words in common word lists		
In lists	64.5	43.7
Not in lists	27.6	16.8
4. Parts of speech		
Prepositions and pronouns	71.6	53.4
Conjunctions and articles	60.5	61.8
Adverbs and verbs	45.3	21.4
Adjectives and nouns	40.8	22.7

As can be seen from Table 10, in both cases the easiest categories of words to replace, judged on the percentage of correct replacements, were 1-2 letter words, 1 syllable words, words in common word lists, and prepositions, pronouns, conjunctions and articles (or functional/structural words).

In both versions there is a slightly higher mean percentage replacement rate for 'more than two syllables' than for 'two syllables'. In both cases the number of words in the 'more than two syllables' subdivision was small - three for pattern 4 and seven for pattern 5. In pattern 4 one of the words - 'aborigines', and in pattern 5 two of the words - 'witchetty' and 'another' were highly redundant in the context and their high replacement rate pushed up the mean. Excluding these three words the mean percentage correct would have been 13.7 instead of 36.2 and 22.4.

Of the fifty words deleted in pattern 4, eleven fell into all four 'easy' categories. There was a 76.9% replacement rate for these words, whilst there was only a 27.6% replacement rate for the seven words that could not be fitted into any of the four 'easy' categories. For pattern 5 the same pattern appeared, with a 59.8% replacement rate for words in all four 'easy' categories and 13.6% for those not in any.

Summary

Four ways of categorising words deleted from passages when using the cloze procedure were determined. Using these as a basis, the percentage errors made by subjects in the study of Clark and Johnson (1972) were computed. These figures showed that for those subjects the easiest words to replace were words that were:-

(a) 1-2 letters
(b) of 1 syllable
(c) that are in common word lists
(d) that are functional/structural words.

All the words for each of the possible seven word deletion patterns

for 350 word passages from Doug of Australia and Deserts were categorized. The 'easy' word categories were used to predict the 'easiest' and 'hardest' patterns for each of these passages. Tests using these patterns were then given to grade 6 children in eight Melbourne metropolitan schools. Mean correct replacement scores for the 'easiest' patterns were significantly higher than those for the 'hardest' patterns. The percentage of correct replacements for the two versions of the two passages supported the chosen categories as being the 'easiest'.

Conclusion

The results of this pilot study seem to indicate that it is quite possible that the difficulty of cloze tests over the same passage using a given n^{th} word deletion can differ significantly, depending upon what particular words are chosen for deletion. For example, if there are approximately 350 words in a passage and every seventh word is deleted, there are seven possible groups of words that can be deleted. Although these are equivalent forms in theory, they are not necessarily equivalent in fact - their difficulty levels might be quite significantly different.

This difference in difficulty levels probably does not matter when results of cloze tests are simply used to rank children. If all the subjects have been treated in the same way the particular deletion pattern probably makes little difference in the rank order. But when the cloze test is for the purpose of obtaining a score which is then to be interpreted, as it is with comparable cloze and multiple-choice scores, or the cloze criterion scores of Bormuth (1971) and that interpretation is based on a single criterion score, then the level of difficulty of the particular cloze deletion pattern used is of importance.

The findings of the pilot study suggest that there is a need to determine ways of overcoming this difficulty. Two possible means of dealing with the problem are:-

(a) some simple method by which the class teacher could adjust the scores on the test to make allowance for the degree of difficulty of the test;

(b) the determination of a range of scores as the criterion for interpreting performance rather than the use of a single score.

The purpose of the main study reported in the next chapter is to investigate these two possibilities.

CHAPTER IV

EXPERIMENTAL DESIGN

As the pilot study had indicated the possibility that any one cloze test from a particular passage might be more, or less, difficult than any other cloze test from the same passage, it was decided to attempt to establish the following:

1. The characteristics of omitted words which influence the difficulty levels of cloze tests.
2. A simple means whereby classroom teachers could adjust the obtained cloze score for any individual, the adjustment to be related to the relative difficulty or ease of replacement of the words deleted in the particular cloze pattern used for the test.
3. An operationally determined cloze criterion score which would indicate whether material was suitable for a child's independent or unsupervised reading. This present investigation concentrated only on the independent level of reading, i.e. the level associated with a 90% minimum performance on a multiple-choice test on the material, because of the large number of cloze tests required to determine this criterion score effectively for any level.

In order to achieve these the following procedures were used:-

1. Using the means of categorizing deleted words developed in the pilot study, the replacement rates for all possible deletion patterns from a large number of passages were determined. In this way 'easy' to replace and 'difficult' to replace categories of words could be determined, and thus the characteristics of omitted words which influence the difficulty level of cloze tests could be established.

2. Using the 'easy' to replace categories determined in (1) above as predictor variables, and the percentage replacement scores as the criterion variable, the number of words in each of the 'easy' categories for each test, together with the percentage replacement score for each test, were entered into a multiple regression analysis. By this means a formula, or formulas, could be established which would allow the teacher to adjust the obtained cloze score for any individual in terms of the numbers of words with certain characteristics in that particular test.

3. By determining the mean replacement score for a large number of cloze tests, involving all the possible deletion patterns of a number of different passages estimated to be at the independent level of reading for the subjects, an operationally determined cloze criterion score could be established which would indicate whether material was suitable for a child's independent or unsupervised reading. The standard deviation associated with this mean score would give a range of scores which, used in conjunction with the mean criterion score, would indicate the efficiency, or relative inefficiency, of a single criterion score.

To meet the needs of the procedures outlined above, each of the possible every seventh word deletion patterns from 16 different 350 word (approx.) passages from books estimated to be at the independent level of reading for the children involved were used. Thus 112 different cloze tests were devised, with 5,600 words deleted.

The Instruments

The cloze tests were devised in the following manner.

(a) Four grade six teachers in Melbourne metropolitan schools were chosen on the recommendations of lecturers from the State College of Victoria at Toorak and school principals. The grounds for the recommendations were that these four were excellent teachers, had taught for at least ten years,

and had an interest in the teaching of reading. Two of the four were the reading co-ordinators for the upper grades in their respective schools, whilst the other two were responsible for co-ordination of all subjects in grades five and six at their schools.

(b) The four teachers were asked to rank the children in their grades on the basis of their reading comprehension ability. Two of the teachers did this on the basis of their knowledge of the children's ability, and as this ranking was done in early December it would be expected that the teachers would know the children well and that the ranking would be reasonably reliable. The other two made their rankings on the basis of their knowledge of the children's performance together with information from a series of reading tests given during the year.

(c) The first 28 children in each grade were then divided into four groups of seven, the first seven in order being designated Group 1, the second seven being designated Group 2, and so on. The members of these groups were judged to be of relatively equal ability although obviously there was some spread, with that spread most likely to be most pronounced in Group 1 (the best readers) and Group 4 (the poorest readers) for each grade. It should be pointed out however that all grades were in excess of 28 and that the poorest readers were not included in the investigation. A cross check with the comprehension test scores for the two grades for which these were available indicated that the ranges, for those two grades at least, were not excessive in any group.

(d) The teachers were then asked to choose one book for each group that they considered would be at the independent level of reading for the children in that group, i.e. they were asked to choose books that the children could be expected to read and comprehend without assistance.

(e) In all cases the book chosen was an anthology of selections. Thus, in each case the passage chosen for the cloze tests was taken from a story chosen at random from those in the book. A list of the books chosen, and the passages used, can be found in Appendix F. In all cases except one,

where the story was little more than 350 words long, the passage chosen allowed a 'run-in' of somewhere between one sentence and one paragraph before deletions commenced, thus allowing some experience of the flavour and tone of the passage.

(f) Seven deletion patterns were then prepared for each passage by deleting words 1, 8, 15, etc., words 2, 9, 16, etc. etc. In determining the 350 words for each passage the following decisions were made. Words such as 'I'll', 'it's' and 'we're' were counted as one word; where words were hyphenated, such as 'co-worker', 'fore-flippers', 'whip-poor-wills' and 'tree-tops', each of the units was treated as a single word; where numbers appeared in the text, e.g. "in the winter of 1774" and "a 64-gun salute", these were each counted as one unit. Thus 1774 and 64 were each counted as single words.

(g) Each of the words was then allocated to its appropriate subdivision of each of the four categories. The categories of number of letters, number of syllables and 'in' common words lists were handled in the same way as in the pilot study. For the part of speech category the part of speech of each of the 5,600 words was determined by the investigator with The Concise Oxford Dictionary as the basic source of reference. Random checks were made of the categorizations by two senior lecturers in English. Instead of using all eight subdivisions previously used (see p.38), only two subdivisions were used, viz. parts of speech found to be 'easy' to replace in the pilot study, and an 'other' group. The 'easy' subdivision was comprised of personal, personal possessive and relative pronouns, prepositions, conjunctions and the definite articles 'the', 'a' and 'an'. All other words were placed in the 'other' or 'hard' to replace subdivision.

The part of speech category was by far the most difficult to use in that many words can be more than one part of speech, depending on the particular usage, and the line between two possible parts of speech is rather fine in some cases. Thus the possibility of placing a word in the wrong subdivision is much greater in this category than it is in any of the others.

Appendix G gives an example of the categorizations. It shows the categorization of the seven patterns from the passage from The Musical Seal.

(h) The tests were prepared by photostating the passages, painting out the words to be deleted with liquid retype, and placing a number, in series, in each of the blanks. A separate answer sheet was provided. Samples of the tests and the answer sheet can be found in Appendix H.

Subjects

The 112 subjects used in this investigation were grade six children from four Melbourne metropolitan State Primary Schools. The total group was almost exactly divided between boys and girls, although there were variations in this relationship from grade to grade and from sub-group to sub-group.

Test Administration.

All testing was carried out by the author under normal classroom conditions.

The seven different patterns for each passage were randomly allocated to the members of each group. After the material had been handed out changes were made to the seating in the room to ensure that no child was sitting next to another doing a test from the same material.

The following instructions were then read, the children following from individual copies:-

"On this page is a reading puzzle. Every seventh word has been left out of a paragraph from a book, and a number has been put where each word was left out. Your job will be to try to solve the puzzle by trying to guess the words left out. You have been given a separate answer sheet to write your answers on. The first answer has been written in to show you what to do.

It will help you in doing this exercise, and the longer one we are also going to do, if you remember these things -

1. Write only one word for each numbered space.

2. Try to fill every blank. Don't be afraid to guess.
3. If you find any very hard, leave that one and come back to it later.
4. Your spelling doesn't matter as long as we can tell what word you meant.
5. In the longer puzzle you might find that there is a number (e.g. 34) or a date (e.g. 1973) missing, rather than a word. "

The subjects then did the following short practice exercise:-

When something hot and something cold 1 brought together, heat will always move 2 the hotter thing to the cooler 3 . Drop some ice cubes in a 4 of warm lemonade. The heat from 5 warm lemonade will go into the 6 cubes. The lemonade will be cooler 7 some of the heat has gone 8 of it. The ice will melt 9 heat has gone into it.

Four minutes were given to complete this practice exercise. The correct answers were then given, followed by a brief discussion of the reasons for certain words being the correct replacement. An opportunity was then given to ask questions. Then the final instructions were given:-

"We are now going to do a much longer puzzle - there are fifty words missing this time. Everyone is doing a different puzzle. You will see that the missing words have been replaced by numbers and that the space will give you some clue as to the length of the missing word.

We are trying to find out how boys and girls like you, in a number of different grades in a number of different schools, can do puzzles like these. Please try your hardest. "

No time limit was set for completion of the tests. The children handed in their sheets as they were completed or satisfied they had replaced as many words as they could. After thirty minutes all remaining tests were collected. In all these latter cases the children had replaced as many words as they could.

Processing the data.

Although 112 tests were prepared and allocated to subjects, the results of only 110 tests are reported. Two subjects in School 4 Group 2 were absent on the day of testing, and as this took place very near to the end of the school year, the school program did not allow for the testing of these children at a later date.

All tests were scored on the basis of one point being given for each correct replacement of a deleted word. Thus, the highest possible score was 50 for any subject. For some purposes the scores have been expressed as percentages. Where this is the case it is clearly indicated in the text.

1. The score for each of the 112 tests was obtained. From these scores the mean and standard deviation for each passage and for all the passages combined was computed. (See Appendix I)

2. The number of words correctly replaced for each subdivision of each of the four categories was determined for each test. These were then summed to determine the percentage replacement rate for each of the subdivisions of each of the four categories for each passage and all passages combined. (See Appendix J)

3. This data was used to determine the easiest subdivisions of each of the four categories.

4. The number of words for each test in each of the easy subdivisions for the four categories - words 1-4 letters long, words of 1 syllable, words of 1-2 syllables, words in common word lists, and words that were either articles, conjunctions, prepositions or pronouns - together with the percentage of correct replacements for each test, was entered into a multiple regression analysis. The computer program, Program Regran (Veldman, 1967) was used for this analysis.

CHAPTER V

ANALYSIS OF THE DATA

Characteristics of omitted words and difficulty of replacement.

All 5,600 words in the 16 passages were categorized according to -

(a) length of word,
(b) number of syllables,
(c) common word lists, and
(d) part of speech, using the subdivisions devised for the pilot study. Mean replacement rates were then computed for each subdivision for each passage and for all passages.

(a) Length of word.

Table 11 shows the number of words in each subdivision of the length of word category, together with the number of these words correctly replaced and the replacement rate for each passage and for all passages combined.

TABLE 11

Replacement of words according to the number of letters.

Passage	Number of letters per word							
	1/2		3/4		5/6		7	
School 1								
1. Musical Seal	75	56	130	78	78	27	67	13
	74.67%		60.00%		34.62%		19.40%	
2. Paul Revere	68	49	131	66	62	19	83	21
	72.05%		50.38%		30.64%		25.30%	
3. Loaded Dog	46	26	162	89	81	19	61	19
	56.52%		54.94%		23.46%		31.15%	
4. Aunt Letty	47	25	187	98	72	22	44	2
	53.19%		52.40%		30.55%		4.54%	

Passage	Number of letters per word							
School 2	1/2		3/4		5/6		7	
1. The Claimant	66	57	140	93	70	33	74	20
	86.36%		66.43%		47.14%		27.03%	
2. The Smiths	71	50	136	61	77	21	65	7
	70.42%		44.85%		27.27%		10.76%	
3. Insight	71	56	132	71	79	21	68	11
	78.87%		53.78%		26.58%		16.17%	
4. Frog Prince	78	60	193	132	60	28	19	9
	76.92%		68.39%		46.66%		47.36%	
School 3								
1. Afghanistan	75	68	132	79	86	34	55	12
	90.66%		59.84%		39.53%		21.81%	
2. Puddin' Thieves	63	44	174	71	53	9	60	14
	69.84%		40.80%		16.98%		23.33%	
3. Paddington Bear	75	61	149	98	69	30	57	18
	81.33%		65.77%		43.47%		31.57%	
4. Jack	69	39	176	97	62	18	43	6
	56.52%		55.11%		29.03%		13.95%	
School 4								
1. Wouldn't Box	86	73	150	102	55	22	59	32
	84.88%		68.00%		40.00%		52.24%	
2. Christmas Trees	48	33	114	57	42	6	46	9
	68.15%		50.00%		14.29%		19.57%	
3. Seal Family	51	30	176	89	79	34	44	7
	58.82%		50.57%		43.04%		15.91%	
4. Rip Van Winkle	72	38	169	52	71	16	38	1
	52.77%		30.76%		14.08%		2.63%	
TOTALS	1061	765	2451	1333	1096	359	883	201
	72.10%		54.38%		32.75%		22.76%	

In all cases the 1-2 letter words were the easiest to replace, with the next easiest being the 3-4 letter words. In five cases (School 1 passage 3,

School 2 passage 4, School 3 passage 4 and School 4 passages 1 and 2) a higher percentage of 7 or more letter words was replaced than for 5-6 letter words. In most of these cases this was probably due to the fact that a number of words were repeated a number of times, and although these words were relatively long, e.g. princess, Stanley, Matthews, and football, they were highly replaceable in the particular contexts.

The total replacement percentages for 1-2 letter words (72.10), 3-4 letter words (54.38), 5-6 letter words (32.75) and 7 or more letter words (22.76), supports the findings of the pilot study regarding the relative ease of replacement according to the number of letters in the deleted word.

For the purposes of the multiple regression analysis referred to later in the chapter, the 1-2 letter and 3-4 letter subdivisions were combined to give the 'easy' subdivision within the category length of word.

(b) Number of syllables

Table 12 shows the number of words in each subdivision of the number of syllables per word category, together with the number of these words correctly replaced, and the replacement rate for each passage and for all passages combined.

TABLE 12

Replacement of words according to the number of syllables

Passage	Number of syllables per word.					
School 1	1		2		3 plus	
1. Musical Seal	236	148	77	24	37	2
	62.71%		31.17%		5.41%	
2. Paul Revere	227	122	88	27	35	8
	53.74%		30.68%		22.86%	
3. Loaded Dog	247	128	83	22	20	3
	51.82%		26.51%		15.00%	
4. Aunt Letty	276	131	66	16	8	0
	47.46%		26.67%		0.00%	

Passage	Number of syllables per word.					
	1		2		3 plus	
School 2						
1. The Claimant	222	156	94	41	34	6
	70.27%		43.62%		17.65%	
2. The Smiths	251	125	70	11	28	3
	49.80%		15.71%		10.71%	
3. Insight	237	139	88	18	25	2
	58.65%		20.45%		8.00%	
4. Frog Prince	291	195	57	33	2	1
	67.01%		57.89%		50.00%	
School 3						
1. Afghanistan	246	165	71	23	31	5
	67.07%		32.39%		16.13%	
2. Puddin' Thieves	249	119	74	17	27	2
	47.79%		22.97%		7.41%	
3. Paddington Bear	252	171	69	26	29	10
	67.86%		37.68%		34.48%	
4. Jack	269	141	70	18	11	1
	52.42%		25.71%		9.09%	
School 4						
1. Wouldn't Box	253	183	72	35	25	11
	72.33%		48.61%		44.00%	
2. Christmas Trees	175	95	43	6	28	4
	54.29%		13.95%		14.29%	
3. Seal Family	264	134	68	17	18	3
	50.76%		25.00%		16.67%	
4. Rip Van Winkle	259	87	75	19	16	1
	33.59%		25.33%		6.25%	
TOTALS	3954	2239	1165	353	374	62
	56.63%		30.30%		16.58%	

In all cases words of one syllable were the easiest to replace, and in only one case (School 4, passage 2) was there a higher percentage replacement for three syllable words than for two syllable words, and even then the difference was negligible (14.29% for three or more syllables, 13.95% for 2 syllables).

In many cases the percentage replacement rate for three or more syllable words was very low, e.g. 0.00% (School 1, passage 4), 5.41% (School 1, passage 1), 6.25% (School 4, passage 4) and 7.41% (School 3, passage 2). For the passage from the Frog Prince the percentage replacement rate was 50%, but only two of the words were three or more syllables long.

The total replacement percentages for one syllable words (56.63), two syllable words (30.30) and three or more syllable words (16.58), support the findings of the pilot study regarding the ease of replacing words according to the number of syllables. For the purposes of the multiple regression analysis the replacement scores for both one syllable words, and one and two syllable words combined, were used as 'easy' subdivisions of the number of syllables per word category.

(c) Words in common word lists.

Table 13 shows the number of words in each of the two subdivisions for this category, together with the number of words correctly replaced, and the replacement rate for each passage and for all passages combined.

TABLE 13

Replacement of words according to whether they were 'in' or 'not in' common word lists.

Passages	In common 200 word lists.		Not in common 200 word lists.	
School 1				
1. Musical Seal	203	137 (67.48%)	147	37 (25.18%)
2. Paul Revere	187	112 (59.89%)	163	45 (27.61%)
3. Loaded Dog	211	115 (54.51%)	139	38 (27.34%)
4. Aunt Letty	226	123 (54.43%)	124	24 (19.36%)

Passages	In common 200 word lists		Not in common 200 word lists	
School 2				
1. The Claimant	208	152	142	51
		73.08%		35.92%
2. The Smiths	204	115	146	24
		56.38%		16.44%
3. Insight	217	137	133	22
		63.14%		16.55%
4. Frog Prince	271	182	79	47
		67.16%		59.50%
School 3				
1. Afghanistan	214	151	136	42
		70.56%		30.89%
2. Puddin' Thieves	223	111	127	27
		49.78%		21.26%
3. Paddington Bear	231	169	119	38
		73.16%		31.94%
4. Jack	235	135	115	25
		57.45%		21.74%
School 4				
1. Wouldn't Box	236	173	114	56
		73.31%		49.13%
2. Christmas Trees	163	92	87	13
		56.45%		14.95%
3. Seal Family	220	122	130	38
		55.46%		29.23%
4. Rip Van Winkle	209	80	141	27
		38.28%		19.15%
TOTALS	3458	2106	2042	554
		60.91%		27.13%

The percentage of 'in' words correctly replaced ranges from a low of 38.28% (School 4, passage 4) to a high of 73.16% (School 3, passage 3), whilst for 'not in' words the rates ranged from a low of 14.95% (School 4, passage 2) to a high of 59.50% (School 2, passage 4).

In all cases the percentage of 'in' words correctly replaced was higher than for 'not in' words. The overall difference of 32.78% in the rates of replacement supports the findings of the pilot study that words 'in' common word lists are much easier to replace.

For the purposes of the multiple regression analysis the replacement scores for 'in' common word lists were used.

(d) Parts of speech.

Table 14 shows the number of words in each of the two subdivisions for this category, together with the number of words correctly replaced, and the replacement rates for each passage and all passages combined. Whereas eight separate subdivisions had been used in the pilot study, a separate subdivision for each part of speech, in this case the words were divided into only two subdivisions, articles, conjunctions, prepositions and pronouns in one and all other parts of speech in an "other" subdivision.

The percentage of words in the articles etc. subdivision correctly replaced ranged from a high of 81.62% (School 2, passage 1) to a low of 47.87% (School 4, passage 4), whilst for the 'other' subdivision the range was from a high of 56.53% (School 2, passage 4) to a low of 21.89% (School 4, passage 4).

In all cases the percentage of articles etc. replaced was higher than for 'other' parts of speech. The overall difference of 29.93% in the rates of replacement supports the findings of the pilot study that words that are articles, conjunctions, prepositions or pronouns are easier to replace on average than are words that are any of the other parts of speech.

For the purpose of the regression analysis the articles etc. subdivision was used as the predictor variable.

TABLE 14

Replacement of words according to part of speech

Passage	Pronoun/Preposition Conjunction/Article		Other	
School 1				
1. Musical Seal	138	97	212	77
		70.29%		36.32%
2. Paul Revere	133	88	217	69
		66.17%		31.80%
3. Loaded Dog	138	76	212	77
		55.08%		36.32%
4. Aunt Ietty	138	81	212	66
		58.70%		31.14%
School 2				
1. The Claimant	136	111	214	92
		81.62%		42.99%
2. The Smiths	136	83	214	56
		61.03%		26.17%
3. Insight	128	88	222	71
		68.75%		31.99%
4. Frog Prince	166	125	184	104
		75.31%		56.53%
School 3				
1. Afghanistan	115	84	235	109
		73.05%		46.39%
2. Puddin' Thieves	138	80	212	58
		57.98%		27.36%
3. Paddington Bear	144	112	206	95
		77.78%		46.12%
4. Jack	133	82	217	78
		61.66%		35.95%

Passage	Pronoun/Preposition Conjunction/Article		Other	
School 4				
1. Wouldn't Box	130	106	220	123
		81.54%		55.91%
2. Christmas Trees	103	68	147	37
		66.02%		25.17%
3. Seal Family	123	76	227	84
		61.79%		37.01%
4. Rip Van Winkle	117	56	233	51
		47.87%		21.89%
TOTALS	2116	1413	3384	1247
		66.78%		36.85%

Regression analysis

The individual data was then processed by means of a multiple regression analysis using the computer program Regran (Veldman, 1967).

The analytic procedure incorporated in the program involves the use of multiple predictors and a single criterion. A set of "beta" weights are then determined for these predictor variables that will produce composite predicted scores which will correlate maximally with the criterion variable.

For the purposes of this present analysis the following 'easy' subdivisions of the four categories were used as predictor variables: words of 1-4 letters (Predictor 1), words of one syllable (Predictor 2), words of 1-2 syllables (Predictor 3), words 'in' common word lists (Predictor 4) and words that were either articles, conjunctions, prepositions or pronouns (Predictor 5).

The input data is shown in Appendix I. For each test the number of words in each of these easy subdivisions is shown under predictor variables,

and the actual score obtained by the subject doing that test is shown under the criterion.

Table 15 shows the correlation matrix for the variables used in this study.

TABLE 15

Correlation Matrix

	Predictor Variables					Criterion
	1-4 letters	1 syllable	1-2 Syllables	In common words	Prep/Pro Conj/Art	%age obtained
1-4 letters		0.7602*	0.4278*	0.7667*	0.5129*	0.1290
1 syllable	0.7602		0.5525*	0.6163*	0.3018*	0.0477
1-2 syllables	0.4278	0.5525		0.3692*	0.0127	0.0125
In common words.	0.7667	0.6163	0.3692		0.5117*	0.3468*
Prep/Pro Conj/Art	0.5129	0.3018	0.0127	0.5117		0.2442**
%age obtained.	0.1290	0.0477	0.0125	0.3468	0.2442	

Significance:

* p .01

** p .05

Amongst the predictor variables the highest correlation (0.7667) was between words 'in' common word lists correct and words 1-4 letters long, with the correlation between one syllable words and words that are 1-4 letters long being only fractionally lower (0.7602). Apart from the correlation between articles, conjunctions etc. and words of one or two syllables, which was only 0.0127, all the correlations were significant.

For the correlations between the predictor variables and the criterion variable, the highest correlation was that of 0.3486 for the number of words

'in' common word lists correct, with the number of articles, conjunctions, etc. correct next best (r = 0.2442). The two predictor variables involving the number of syllables showed very low correlations with the criterion (0.0477 and 0.0125).

Table 16 shows the cumulative variance for the predictor variables in combination, commencing with the best single predictor. The order of adding in of predictor variables was determined by the computer.

TABLE 16

Cumulative Variance for the best combinations of predictor variables

Predictor Variables	Cumulative
4 (common words)	0.1203
1 (1-4 letters)	0.1658
5 (Preps etc.)	0.1786
2 (1 syllable words)	0.1888
1	0.1892
3	0.1894
1	0.1895
2	0.1896
1	0.1896
2	0.1897
1	0.1897
2	0.1897

An examination of the iteration sequence shows that 12.03% of the variance in the total percentage scores is accounted for by the number of words 'in' common words lists (Predictor 4), and that this is the best single predictor. The addition of number of words of 1-4 letters correct (Predictor 1) adds another 4.55%, whilst the addition of the number of articles, conjunctions, etc. correct (Predictor 5) adds a further 1.28%.

As a result, five possible predictor models were determined, viz:

Model 1 Predictor 4
Model 2 Predictors 1 and 4
Model 3 Predictors 1, 4 and 5
Model 4 Predictors 1, 2, 4 and 5
Model 5 All predictors,

with Model 1 being the best single predictor, and Model 2 the best pair of predictors.

Table 17 shows the correlation, variance, Beta and B weights, and the regression constant for each of these five models.

TABLE 17
Correlation, Variance, Beta, B and Regression Constant scores for five models.

Model	Predictors	r	r^2	Beta	B	Reg. Const.
1	4	0.3468	0.1203	0.3468	1.2890	7.6240
2	1	0.4072	0.1658	-0.3321	-1.2066	16.3955
	4			0.6015	2.2354	
3	1	0.4266	0.1820	-0.3766	-1.3681	15.1948
	4			0.5579	2.0734	
	5			0.1518	0.5953	
4	1	0.4354	0.1896	-0.2887	-1.0487	23.1296
	2			-0.1268	-0.5283	
	4			0.5756	2.1393	
	5			0.1360	0.5332	
5	1	0.4355	0.1897	-0.2812	-1.0214	26.8868
	2			-0.1254	-0.5225	
	3			-0.0157	-0.0998	
	4			0.5786	2.1504	
	5			0.1288	0.5051	

The Beta weights, which are standard partial regression weights, indicate the extent to which each variable is utilized in the regression equation, whilst the B-weight vector, with the regression constant added, gives the information scaled in terms of the raw scores of the predictor variables. (Veldman, 1967).

The predicted percentage scores for each subject for each model were then computed, together with the adjusted percentage scores. An example is given.

Example from Model 2 (Predictors 1 and 4)

Subject	Pred 1	Pred 2	Pred 3	Pred 4	Pred 5	Obtained percent
0101	28.000	30.000	45.000	26.000	21.000	52.000

Predicted percentage score = $B_1X_1 + B_4X_4$ + Regression Constant

= (-0.3321 x 28.000 + 0.6015 x 26.000) + 16.3955

= 40.732

Adjusted percentage score = Obtained criterion score plus the difference between the predicted percentage score and the mean criterion percentage score.

= 52.000 + (-40.732 + 48.1455)

= 59.4135

A series of F tests was then carried out. Two of these were concerned with the significance of the prediction obtained by using (a) all predictors, and (b) the best single predictor (Predictor 4). Both were significant at the .001 level.

The remaining F tests were carried out in order to examine the predictive efficiency gained by adding predictors to the equation. Only one of these, adding Predictor 1 to the best single predictor (Predictor 4), led to significant improvement.

Table 18 summarizes the results of the F tests.

TABLE 18

F Test Results

Predictors	D.F.	F ratio	p
All	5/104	4.870	0.0007
4	1/108	14.769	0.0004
4 + 1 vs 4	1/107	5.833	0.0165
4 + 1 + 5 vs 4 + 1	1/106	2.102	0.1462
4 + 1 + 5 + 2 vs 4 + 1 + 5	1/105	0.982	0.6752
4 + 1 + 5 + 2 + 3 vs 4 + 1 + 5 + 2	1/104	0.016	0.8958

This clearly indicates that Predictor 4 alone is almost as profitable as using all five predictors although a significant increase is obtained by adding Predictor 1. However no profit is achieved by adding any of the other predictors.

As the purpose of this section of the investigation was to choose a simple means of adjusting the scores to make allowance for the characteristics of the words deleted in the particular cloze pattern, the results seem to indicate two possibilities. The first is to use the best single predictor (Predictor 4 - the number of words in common word lists), whilst the second is to use the most efficient pair of predictors (Predictor 4 plus Predictor 1 - the number of words 1-4 letters long).

For teachers to use these predictors to adjust obtained cloze scores the following procedures would be required.

1. Using Predictor 4 alone.

   (a) Determine the number of deleted words in the passage that appear in the composite common words list.

(b) Compute the predicted score by multiplying the number of deleted words in the common words list by 1.2890 and add 7.6240.

(c) The adjusted score would then be the actual total replacement score obtained by the child plus the difference between the predicted percentage score and the mean criterion score.

2. Using Predictors 4 and 1

(a) Determine the number of deleted words in the passage that appear in the composite common words list and the number of deleted words that are 1-4 letters in length.

(b) Compute the predicted percentage score as follows:
Predicted score = (-1.2066 times the number of words 1-4 letters long plus 2.2354 times the number of words in common word lists) plus 16.3955.

(c) Compute the adjusted percentage score by the actual total replacement score obtained by the child plus the difference between the predicted percentage score and the mean criterion score.

Using Predictor 4 alone would be reasonably simple and would not require very much work on the part of the classroom teacher. Using Predictors 4 and 1 would only be relatively more time consuming and difficult to use. The decision as to whether to recommend their use however is dependent on a number of points.

(a) Predictor 4 accounts for only 12.03% of the total variance, and Predictors 4 and 1 together account for only 16.58%. Thus, although the formulas that arise out of the weights found for the predictors in this investigation are relatively simple, the predictors probably account for too little variance to warrant recommending the use of these formula to adjust the obtained scores.

(b) The argument often voiced against readability formulas that they are too 'mathematical' for teachers to be bothered to use, could

easily apply in this case. The mathematics required in the formulas arising out of this present study is similar to that required in the readability formulas mentioned in Chapter 1.

(c) It must be remembered that in the case of the 'strongest' predictor (Predictor 4) the data has been based on the number of words in a composite common words list. The use of the adjusting procedure arising out of this present study would require the teacher to have this particular list on hand. Of the four word categories and the five predictor variables used in this study this is the only one that requires the teacher to have any special information. It is therefore less likely that teachers would make use of this information than had the best predictor(s) been the number of 1-4 letter words and/or the number of one syllable words, both of which are very simple to determine and neither of which require further reference to any other information.

(d) The effectiveness of the use of the weights can be shown by comparing the standard deviations for the obtained and adjusted percentage scores in this present study.

Table 19 shows the means and standard deviations for the obtained scores and the scores when adjusted by the use of the weights for Predictor 4 and for Predictors 1 and 4.

TABLE 19

Means and standard deviations for obtained and adjusted scores

		Obtained percentage	Adjusted percentage
Predictor 4	Mean	48.145	48.145
	S.D.	14.776	13.859
Predictors 4 and 1			
	Mean	48.145	48.145
	S.D.	14.776	13.496

An analysis of the results shown in Table 19 indicates that for each of these the difference between the standard deviation for the obtained percentage score and the adjusted percentage score, although giving some advantage, is so small as to hardly justify the work involved in using the weights to adjust the scores.

If the best single predictor is used the difference in standard deviations is only 0.917, whilst for the best pair of predictors the difference is 1.280. There appears therefore that there is little justification in expecting teachers to go to all the work involved in making the adjustments when overall there is only a small difference in the standard deviation.

Thus it appears as if this investigation has not succeeded in devising a simple practical means for adjusting obtained cloze scores in terms of the characteristics of the deleted words that makes sufficient difference to warrant its general acceptance.

Operationally defined cloze criterion score.

The third aspect of the investigation was to determine operationally a cloze criterion score for performance on material estimated to be suitable for children's independent or unsupervised reading.

Table 20 shows the individual scores for each deletion pattern and the mean and standard deviation for each passage.

TABLE 20

Individual scores for each deletion pattern and mean and standard deviation for each passage.

Passage	Deletion Pattern 1	2	3	4	5	6	7	Mean	S.D.
School 1									
Musical Seal	26	22	26	25	20	32	23	24.85	3.56
Paul Revere	23	19	16	29	28	22	20	22.42	4.37
Loaded Dog	21	19	22	14	18	20	39	21.85	7.39
Aunt Letty	8	25	13	22	11	34	34	21.00	9.91
School 2									
The Claimant	28	38	25	24	32	29	27	29.00	4.40
The Smiths	24	22	12	22	16	25	18	19.85	4.36
Insight	16	26	22	26	26	20	23	22.71	3.49
Frog Prince	29	29	31	34	42	28	36	32.71	4.65
School 3									
Afghanistan	31	29	23	29	29	24	28	27.57	2.72
Puddin'Thieves	23	18	23	18	21	16	19	19.71	2.49
Paddington B.	31	30	27	38	28	25	28	29.57	3.89
Jack	30	21	28	21	21	19	20	22.86	3.98
School 4									
Wouldn't Box	29	36	36	25	34	36	33	32.71	3.92
Christmas Ts.	26	22	12	20	*	25	*	21.00	4.98
Seal Family	24	18	28	43	7	19	21	22.86	10.19
Rip Van Winkle	16	24	15	22	8	9	13	15.29	5.60

* These two were not attempted due to the absence of the subjects.

An analysis of the individual scores shows a substantial range, with a low of 7 and a high of 43, both of which occurred in the same passage (School 4, passage 2).

The mean scores for the passages range from a low of 15.29 (School 4, passage 4) to a high of 32.71 (School 2, passage 4, and School 4, passage 1).

The standard deviations for these passages are comparatively low and therefore it could be suggested that Rip Van Winkle was, relatively, a poor choice being too hard (not one of the seven subjects had a replacement rate of more than 48%), and Frog Prince and Wouldn't Box, relatively, too easy (all subjects having a replacement rate in excess of 50%).

Two of the passages showed relatively high standard deviations. For School 4, passage 3, the standard deviation was 10.19. This can be explained by the fact that two of the scores, 7 and 43, were grossly different, whereas the rest of the scores were very similar. An investigation of the deletions for these two patterns indicates that they were the easiest and most difficult for the passage, although the difference probably wouldn't account for the big difference in correct replacements. For School 1, passage 4, the standard deviation was 9.91. Apart from the fact that pattern 1 was by far the most difficult - the subject only obtained a score of 8 - the marked variations in scores for this passage can probably only be accounted for by lack of homogeneity in the group.

The overall mean replacement score of 24.07 gives an operationally defined score for cloze tests used in this investigation of 48.145% with a standard deviation of 14.776. If these two figures are rounded off the results indicate a mean of 48% with a range of 33% - 63% (48 ± 15) accounting for two thirds of the scores.

Table 21 compares the results of this study with those of the earlier studies reported in Chapter 2.

TABLE 21

Cloze criterion scores for the Independent level.

Bormuth 1967	50%
Bormuth 1968	57%
Rankin and Culhane 1969	61%
Anderson and Hunt 1972	53%
Boyce 1974	48%

Table 21 indicates that the score operationally obtained in this study is lower than the equivalent scores found in the previous studies, especially that of Rankin and Culhane.

This present study has a standard deviation associated with it however, whereas all the others provided single criterion scores. If it could be assumed that a similar standard deviation could be associated with the criterion scores previously reported, there is a range of scores from 46% to 63% that is common to all students. (See Figure 2).

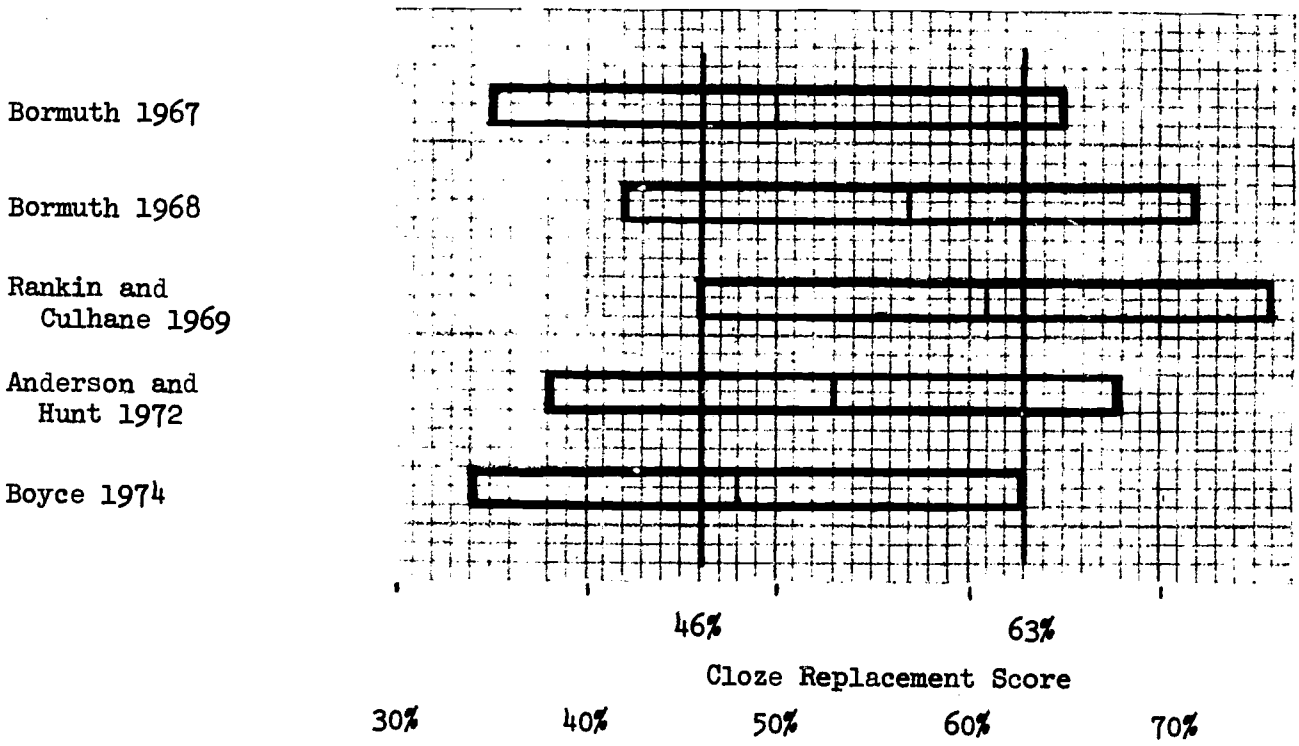

Figure 2. Common range of scores for independent level.

Furthermore, if the standard deviation found in this present study could be associated with the criterion scores for both the independent and instructional levels reported in earlier studies this would show whether it could be expected that any scores would fall into both of these levels.

Figure 3 shows the information for all four previously reported studies showing the criterion scores obtained, together with the range associated with a standard deviation of 15, and the range for each study that is common to both reading levels. Table 22 summarizes this information.

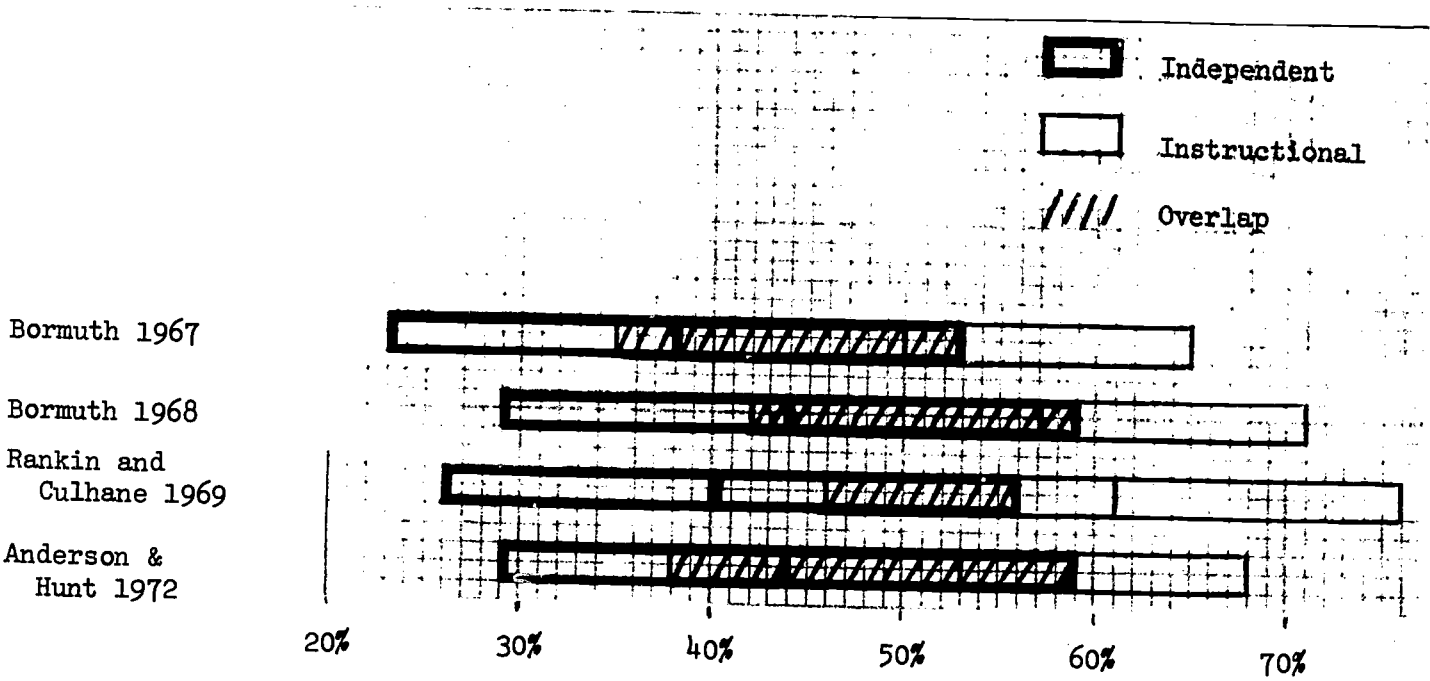

Figure 3. Range of scores associated with both the independent and instructional reading levels (See also Table 22)/

TABLE 22

Range of scores associated with both the independent and instructional levels of reading.

Bormuth 1967	35 - 54
Bormuth 1968	42 - 59
Rankin and Culhane 1969	46 - 56
Anderson and Hunt 1972	38 - 59

In all cases the range of scores that might be expected to be associated with both levels of reading is large - the highest being 22 for Anderson and Hunt and the lowest being 11 for Rankin and Culhane. Whilst it is artificial

to associate the standard deviation from one study with the results of totally different studies, and whilst it may be said that the standard deviation in this present study might be rather high because of the material used and the possible lack of homogeneity in the members of the groups doing the tests, it is still likely that there would be a relatively large range of scores associated with both the independent and instructional levels of reading.

Because of this it seems likely that the use of cloze scores associated simply with instructional and independent levels of reading may lead to rather gross judgements. It is probable that what is needed is a greater degree of differentiation in the material used to obtain appropriate cloze performance levels. For example, it would probably be more appropriate to obtain scores on a variety of levels such as material that is:

(a) much too difficult,
(b) rather difficult but with very high interest,
(c) easy independent level judged by teacher,
(d) independent level judged by child,
(e) much too easy,
(f) instructional level judged by teacher,

and attempt to aim at the maximization of differences between the levels.

CHAPTER VI

SUMMARY AND CONCLUSIONS

This investigation has sought to examine the measurement of passage performance by the use of the objectively determined cloze procedure.

Previous attempts to measure passage performance using this procedure have involved the acceptance of the Kilgallon (1942) - Betts (1946) criteria for frustrational, instructional and independent levels of reading, and have determined cloze scores comparable to the multiple-choice criteria for these levels (Bormuth 1967, 1968; Rankin and Culhane 1969; Anderson and Hunt 1972). One attempt has been made (Bormuth 1971) to determine criteria for passage performance using cloze criteria alone.

All these previous studies have resulted in single cloze scores comparable to the 75% and 90% multiple-choice criteria, or in the case of Bormuth (1971), single cloze criteria scores for optimal efficiency according to the type of reading material and the grade level.

This present study has attempted to inquire whether a single cloze criterion score can be misleading if it is being used to determine the suitability of reading material for an individual, as the score a child obtains appears to be a function of the types of words deleted. It is therefore feasible that for the possible cloze tests over any given passage there may be widely fluctuating levels of difficulty depending upon the actual combination of words being deleted in any one cloze deletion pattern.

Using material said to be at the independent (unsupervised) level of reading for the subjects involved three matters were investigated.

1. The characteristics of deleted words which make them easy or difficult to replace. It was found that the easiest words to replace were those that were 1-2 letters long, were one syllable long, were in common word lists and were articles, conjunctions, prepositions or pronouns.

2. A means of adjusting the obtained scores to make allowance for the characteristics of the words deleted.

For this purpose a regression analysis was used to determine formulas to adjust the actual score obtained by taking into account the difference between the score that would be predicted from the number of deleted words with certain characteristics and the mean criterion score.

Two formulas were determined for computing the predicted score:

(a) using the best single predictor (the number of words in common word lists), and

(b) using the best pair of predictors (the number of words in common word lists and the number of words 1-4 letters long).

Although the formulas derived would have been relatively easy for classroom teachers to use it was decided that the work associated with adjusting the scores was not justified in terms of the actual gain.

Cooley (1971), quoting Burket 1964, Herzberg 1969, and Marks 1966, suggests that predictor weights often do not correlate as well with the criteria for new samples as they did with the original sample. He feels that in many cases "rather simple alternatives to regression weights, such as using the elements of r_c directly (or even just unit weights!) frequently outperform the B weights on cross validation." (p.619) He claims that the problem diminishes as the number of predictors gets larger, at least 10 or 20 to 1. In this investigation only five predictor variables were used.

3. The determination of an operationally obtained cloze criterion score that could be used to determine whether materials is suitable for

the child's unsupervised or independent reading.

For this purpose the overall mean and standard deviation were obtained from 112 different cloze tests (seven different patterns of 16 different passages) used in this investigation. As the passages used for the cloze tests were from books deemed by the teacher to be suitable for unsupervised reading by the children involved, it should follow that the range given by the mean plus or minus one standard deviation should give an estimate of scores that could be expected to be achieved by two thirds of the children doing cloze tests from material suitable for independent reading.

An analysis of the results obtained, together with comparisons made with the results in the earlier reported studies, suggests that there is likely to be an overlap between expected scores on the independent and instructional levels.

The data gathered in this present study does highlight the weakness of a single criterion score as an indication of passage performance. Far more flexibility is required than is given by the oversimplification of all the factors implied in the use of a single criterion.

Some limitations of the study.

There are a number of factors that need to be considered in relation to the results of this investigation.

In the first place the results are only really generalizable in terms of situations where:-

(a) the subjects are sixth grade children;
(b) there is a fifty item, one in seven deletion pattern;
(c) the format is a photocopy of the original passage with the deleted words whited out, and a separate answer sheet provided;
(d) the children do a practice example first;
(e) the same type of instructions are used.

Generalizing the results outside this fairly rigid set of constraints would not be really justified. There is insufficient evidence in the literature regarding the equivalence of cloze performance at various grade levels, using different n^{th} word deletions over the same material, or using different test formats, instructions and strategies. So, for example, it could not be said that the results obtained in this study would be applicable to an every fifth word deletion, using a typed format where the children write the answer in, where the children are instructed to read through the passage to grasp meaning before attempting to replace any words, and where the subjects are fifth or fourth grade children. In fact, a change in any one of these factors might well affect the valid use of the results from this study. There appears to be great scope for a large number of different studies to explore the relationships between factors such as these and the obtained scores.

Secondly, there is the possibility that the materials used for the tests in the main investigation were not uniformly at the independent level of reading for all the children involved. Although the teachers were asked to supply books to meet this criterion, it is possible, or even probable, that the books they chose were directed more to the mean for each group of seven rather than for the group as a whole. This would probably have had little effect, if each group was fairly homogeneous in ability, but it could be argued, particularly for each group 1 (the best readers) and each group 4 (the poorest readers), that the range of ability may have been quite wide. Thus, although this may have balanced out and the effect on the obtained mean score been small, it may have resulted in a larger standard deviation. However, as indicated earlier, the poorest readers in each grade were not included, and a check on the groups in the grades where teachers did have reading test scores available indicates that the groups were relatively homogeneous.

It may have been better, though not very practical, to have chosen a book specifically for each child, and had each child do - at reasonable

intervals - all seven patterns for the chosen passage. Another alternative may have been to have ranked all the children in the four grades according to a standardized test of reading comprehension such as the Schonell R4, the Neale Analysis of Reading Ability, or Daniels and Diaks Test 10. Groups of seven could then have been determined over all four grades by matching scores as closely as possible. The choice of book for each group would then have required the concensus of opinion of the four teachers involved.

A more serious problem is the fact that all the books chosen were in fact anthologies or collections of stories. Because of this the possibility exists that there were wide fluctuations in the difficulty levels of the various stories contained in any one book. Although the passage used from any one book was chosen at random, and this randomizing process may have evened out the possible differences in difficulty, the possibility still exists that some of the passages chosen were not representative of the overall difficulty of the book. It is probable that it would have been better to have used more than one passage from the book, to have used, for example, three passages each 120 words long rather than only one 350 word passage. Possibly an even better solution would have been to have asked the teachers to choose a passage from a book rather than simply a book. In this way there would be more certainty in the belief that the passages chosen were, in the teachers' opinions, representative of the independent level of reading for the children involved.

Finally, in relation to this second problem is the question raised earlier of the reliability of teacher estimates of the suitability of materials. As was quoted earlier, Klare (1963, p.81) states that "they are recognized as subject to considerable error". The teachers chosen for this study were all very experienced teachers, with excellent records as teachers, and with a particular interest in reading. All had taught the children used as subjects in this investigation for twelve months and should therefore have had a good understanding of the ability of each individual in the area of reading, and more specifically, reading

comprehension. Notwithstanding this, however, there must be some doubt as to the aptness of the choice of books made for the particular subjects involved, although it is hoped that the particular choice of teachers has minimized this doubt.

A third factor that needs to be considered is the method used for scoring the responses. The rationale for the exact replacement method used in this investigation has been dealt with earlier in Chapter 2. There must be some doubt however as to whether this is the most appropriate method of scoring to use for this particular purpose. Whilst it is acknowledged that it is the only truly objective method of scoring, and that it is far simpler, the fact that the test is being used to determine whether the material is suitable for an individual must be considered. It is possible that there are a number of different types of replacement that might serve as indicators that the individual is understanding or comprehending the material in the passage. These different types of replacements include not only similes and logical replacements, but also the use of basically the correct word but the wrong number (e.g. man instead of men) or the wrong tense (e.g. runs instead of ran). It is also possible that on some occasions the use of the correct part of speech, even if the word is quite wrong, may indicate a grasp of the material. The concepts of restricted and elaborated codes of language may well be pertinent to this question. It is possible that restricted language users might comprehend the language in the passage but that their own usage might prevent them from replacing certain words correctly, thus leading to artificially low estimates of their comprehension of the material. This could be a possibility for children from lower class or ethnic minority backgrounds. It is also possible that elaborated code users may use enriched vocabulary in some cases, only to find that these are scored as incorrect. The recent work of Poole (1973) is very relevant to this question of social class differences in language and the cloze procedure.

It is therefore possible that one of the major claims to simplicity in the cloze procedure, the scoring of only exact replacements, may work to the disadvantage of some children.

Fourthly, there appears to have been no study of the effects of the cloze procedure on the motivation of the children to succeed.

An investigation of the results in the Pilot Study (Chapter 3) indicates that for the first 20% of deletions, the subjects doing pattern 4 (easy) had a 57.0% correct replacement rate, whereas for those doing pattern 5 (hard) the rate was only 31.1%. From that point on the subjects doing pattern 5 actually did better as the final correct replacement rates were 50.80% for pattern 4 (down 6.2%) and 32.78% for pattern 5 (up 1.68%).

Although the opposite appeared to happen in this case, it could be speculated that if the first part of a cloze test contains a high number of difficult deletions this might have a depressing effect on the subjects. This could be tested by having half the subjects do the 'easy' version of a cloze test and the other half do the first 20% of the hard version before adjusting it to make the final 80% the 'easy' version. The mean scores for the final 80% could then be compared.

It is possible that difficult to replace words at the beginning of a passage do not matter too much if the subjects see their task simply as one of seeing how many gaps they can fill in, thus approaching the task from a 'bit' rather than a 'whole' approach. The usual instruction "You may skip hard blanks and come back to them again" may well reinforce a strategy of not being very concerned about the total context. As was mentioned earlier (Chapter 2) Boyce (1972) found that many subjects filled in blanks with words which were logical in the immediate context, but which were incorrect or even illogical in the total context of the passage. This seems to suggest a 'bit' approach.

Conclusion

This thesis has explored the use of the cloze procedure as a means of determining, by direct testing, the suitability of written material for the individual, and in particular, for his independent, unsupervised reading.

The strengths of the cloze procedure for this purpose lie in its basic simplicity, its objectivity, and its ability to match the child's performance with the actual material. Its basic weakness lies in problems associated with the meaningful interpretation of the score achieved by an individual as a result of cloze tests on the material.

Attempts have been made (Bormuth, 1967, 1968; Rankin and Culhane, 1969; Anderson and Hunt, 1972) to relate scores on cloze tests to scores on multiple-choice tests of the same material and then use these comparable scores as criteria for interpreting the individual's performance, thus determining the suitability of the material for him. This thesis has explored this approach and has shown that there are problems associated with it, problems which cast some doubt on the effectiveness of the cloze procedure for this purpose.

It should be noted however that these doubts about the effectiveness of the cloze procedure apply only to the interpretation of cloze scores for passage performance purposes. The findings of the investigations reported in the thesis do not imply criticism of the procedure for many of the other purposes for which it is used. For many purposes the cloze procedure appears to be an exceptionally robust and useful measure.

BIBLIOGRAPHY

AINLEY, J.G. Changes in the provision of teacher education since 1960. V.I.E.R. Bulletin. 1972a, 29, 16-31.

AINLEY, J.G. The institutional provision for the education of intending teachers: Canada and Australia. Unpublished M.Ed. thesis. University of Melbourne, 1972b.

ANDERSON, J. A scale to measure the reading difficulty of children's books. University of Queensland Papers in Education. 1967, 1, 6, (whole issue).

ANDERSON, J. Choosing the right book. Papua and New Guinea Journal of Education. 1968, 5, 3-6.

ANDERSON, J. Application of cloze procedure to English learned as a foreign language. Unpublished doctoral dissertation. University of New England, 1969.

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

ANDERSON, J. Selecting a suitable 'reader': procedures for teachers to assess language difficulty. RELC Journal. 1971b, 2, 2, 35-42.

ANDERSON, J. Personal communication. 4th December, 1972.

ANDERSON J. and HUNT, A.H. A frame of reference for cloze tests of readability of English learned as a foreign language. Papua New Guinea Journal of Education. 1972, 8, 3, 184-188.

BALL, I.L. and WILLIAMSON, H.J. How readable is literature for children. V.I.E.R. Bulletin. 1973, 31, 14-20.

BETTS, E. A. Foundations of Reading Instruction. 1946, American Book Company, New York.

BICKLEY, A.C., ELLINGTON, B.J. and BICKLEY, R.T. The cloze procedure: a conspectus. Journal of Reading Behaviour. 1970, 2, 3, 232-249.

BIRD, B. and FALK, I. Trend: Teachers Handbook. 1971, Cheshire, Melbourne.

BLAIR, A.M. Everything you always wanted to know about readability but were afraid to ask. Elementary English. 1971, 48, 5, 442-443.

BOND, G. and TINKER, M. Reading Difficulties: Their Diagnosis and Correction. 1967, Appleton Century Crofts, New York.

BORMUTH, J.R. Experimental applications of cloze tests. In Figurel, J.A.(ed.) Improvement of Reading through Classroom Practice. 1964, International Reading Conference, Newark, Del.

BORMUTH, J.R. Validities of grammatical and semantic classifications of cloze test scores. In Figurel, J.A. (ed.) Reading and Inquiry. 1965a International Reading Association, Newark, Del.

BORMUTH, J.R. Optimum sample size and cloze test length in readability measurement. Journal of Educational Measurement. 1965b, 2, 1, 111-116.

BORMUTH, J.R. Readability: A new approach. Reading Research Quarterly. 1966, 1, 3, 79-131.

BORMUTH, J.R. Comparable cloze and multiple-choice test scores. Journal of Reading. 1967, 10, 291-299.

BORMUTH, J.R. Cloze test readability: criterion reference scores. Journal of Educational Measurement. 1968, 5, 189-196.

BORMUTH, J.R. Development of standards of readability: towards a rational criterion of passage performance. Research in Education. 1971, microfiche ED 054 233.

BORTNICK, R. and LOPARDO G.S. An instructional application of the cloze procedure. Journal of Reading. 1973, 16, 4, 296-300.

BOYCE, M.W. An investigation of some aspects of the cloze procedure as used for measuring reading comprehension. Unpublished paper. University of Melbourne. 1972.

BOYCE, M.W. A Comprehensive Bibliography of the Cloze Procedure. 1973a Toorak Teachers College.

BOYCE, M.W. An introduction to the cloze. V.I.E.R. Bulletin. 1973b, 30, 27-34.

BOWERS, F. and NACKE, P.L. Cloze, transformational theory and redundancy. Journal of Reading Behaviour. 1971-72, 4, 1, 20-33.

BURKET, G.R. A study of reduced rank models for multiple prediction. Psychometric Monographs. 1964, No.12.

CAROZZI, B. Studies in remedial teaching. 1972. Unpublished paper, University of Melbourne.

CAVANNA, B. Doug of Australia. 1965. Chatto and Windus.

CHOMSKY, N. Syntactic Structures. 1957. Mouton, The Hague.

CHOMSKY, N. Aspects of the Theory of Syntax. 1965. M.I.T. Press, Cambridge, Mass.

CLARK, M.L. and JOHNSON, B. "Cloze" procedures in measuring comprehension: simple or complex? 1972. Mimeographed paper, A.C.E.R.

CLARK, M.L. and JOHNSON, B. "Cloze" procedures in measuring comprehension: simple or complex? Q.I.E.R. Journal. 1973, 2, 13-22.

COOLEY, W.C. Techniques for considering multiple measurements. In Thorndike, R.L. Educational Measurement. Second edition 1971. American Council on Education. Pp.601-622.

COLEMAN, E.B. Developing a technology of written instruction: some determiners of the complexity of prose. (Paper presented at a meeting co-sponsored by the A.M.E.R. and the I.R.A., Seattle, May 4, 1967).

CULHANE, J. W. Cloze procedures and comprehension. The Reading Teacher. 1970, 23, 5, 410-413.

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

DARNELL, D.K. Clozentropy: a procedure for testing English language proficiency of foreign students. Speech Monographs. 1970, 37, 1, 36-46.

DELLA-PIANA, G. Reading Diagnosis and Prescription. 1968. Holt Reinhardt and Winston.

EDWARDS, R.P.A. and GIBBON, V. Words Your Children Use. 1964. Burke, London.

Education Gazette and Teachers Aid. 1972. 72, 10A, 513-515. Education Department of Victoria.

FLESCH, R. Estimating the comprehension difficulty of magazine articles. Journal of General Psychology. 1943, 28, 63-80.

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

FLESCH, R. Measuring the level of abstraction. Journal of Applied Psychology. 1950, 34, 384-390.

FORBES, F.W. A new method for determining the readability of standardized tests used in counselling. 1953. Unpublished doctoral dissertation, Kansas University.

FROESE, V. Cloze Readability versus the Dale-Chall formula. Research in Education. 1971. ERIC microfiche ED 051975.

FRY, E.B. Readability formula that saves time. Journal of Reading. 1968a, 513-6, 575-8.

FRY, E.B. Developing a word list for remedial reading. In Schell, L.M. and Burns, P.C. Remedial Reading: An Anthology of Sources. 1968b, Allyn and Bacon, Boston.

FRY, E.B. Reading Instruction for Classroom and Clinic. 1972. McGraw Hill, New York.

GALLOWAY, P. How secondary students and teachers read textbooks. Journal of Reading. 1973, 16, 3, 216-219.

GOETZ, D. Deserts. 1956. Wheaton, Exeter.

GRACE, G.R. The changing role of the teacher: implications for recruitment. Education for Teaching. 1967, 72, 51-57.

GRAY, W.S. Gray Oral Reading Tests. 1963. Bobbs-Merrill, Indianopolis.

GRAY, W.S. and LEARY, B.A. What Makes a Book Readable? 1935. University of Chicago Press, Chicago.

GREENE, F.P. A modified cloze procedure for assessing adult reading comprehension. Dissertation Abstracts, 1965, 25, 10, 5734A.

GROFF, P. (ed.) Research critiques. Elementary English. 1971, 6, 675-681.

GUICE, B. The use of the cloze procedure for improving reading comprehension of college students. Journal of Reading Behaviour. 1969, 1, 3, 81-92.

GUSZAK, F.J. A comparative study of the validity of the cloze test and the Metropolitan Achievement Test (Reading comprehension sub-test) for making judgements of instructional levels. Research in Education. 1970, ERIC microfiche, ED 039108.

HARRIS, A.J. Effective Teaching of Reading. 1962. McKay, New York.

HERZBERG, P.A. The parameters of cross-validation. Psychometrika (Monograph Supplement 16). 1969, 34.

HUMPHREYS, J. and KAY, R. The use of the Cloze procedure to grade texts in the Pacific Horizons' reading scheme. Papua and New Guinea Journal of Education. 1971. 7, 2, 5-8.

HUNNICUTT, C.W. and IVERSON, W.J. (eds). Research in the Three R's. 1958. Harper, New York.

HUNT, L.C. The effect of self-selection, interest and motivation upon independent, instructional and frustrational levels of reading. (Paper presented at the 14th Annual Conference of the I.R.A., Kansas City, Missouri, May 1 and 2, 1969.)

HUUS, H. Innovations in reading instruction: at later levels. In Robinson, H.M. (ed.) Innovation and Change in Reading Instruction. 1968. Sixty-seventh Yearbook of the N.S.S.E. Part 11, pp. 126-158.

JENKINSON, M.E. Selected processes and difficulties in reading comprehension. Unpublished doctoral thesis. 1957, University of Chicago.

JOHNSON, C. Context between cloze deletions. 1968. Unpublished paper. University of New England.

JONGSMA, E.R. The cloze procedure: a survey of the research. Research in Education. 1971. ERIC microfiche, ED 050893.

KERR, A.H. Cloze procedure - a 'new' tool for measuring aspects of language communication. Armidale Teachers College Bulletin. 1970, 16, 1, 26-35.

KERR, A.H. and SMITH, A.G. Cloze procedure: standard or exact-length blanks: a methodological problem. 1968. Unpublished paper, University of New England.

KILLGALLON, P.A. A study of the relationships among certain pupil adjustments in language situations. 1942. Unpublished doctoral dissertation, Pennsylvania State College.

KLARE, G.R. Measures of readability of written communication: an evaluation. Journal of Educational Psychology. 1952, 43, 385-399.

KLARE, G.R. The Measurement of Readability. 1963. Iowa University.

KLARE, G.R. Comments on Bormuth's readability: a new approach. Reading Research Quarterly. 1966, 1, 4, 119-125.

KLARE, G.R., SINAIKO, H.W. and STOLUROW, L.M. The cloze procedure: a convenient readability tests for training materials and translations. International Review of Applied Psychology. 1972, 21, 2, 77-103.

KUCERA, H. and FRANCIS, W.N. Computational Analysis of Present Day American English. 1967. Brown University Press, Providence, R.I.

LORD, F.M. Sampling fluctuations resulting from the sampling of test items. Psychometrika. 1955, 20, 1-22.

LORGE, I.D. Predicting readability. Teachers College Record. 1944, 45, 404-419.

LORGE, I.D. Lorge and Flesch readability formulae: correction. School and Society. 1948, 67, 141-142.

LOUTHAN, V. Some systematic grammatical deletions and their effects on reading comprehension. English Journal. 1965, 54, 294-299.

MACGINITIE, W.H. Contextual constraint in English prose paragraphs. Journal of Psychology. 1961, 51, 121-130.

MACGINITIE, W.H. Comments on Professor Coleman's paper. (Paper presented at the Symposium on Verbal Learning Research and the Technology of Written Instruction, Columbia University, 1966).

MCLEOD, J. The estimation of readability of books of low difficulty. British Journal of Educational Psychology. 1962, 32, 2, 112-118.

MCNALLY, J. and MURRAY, W. Key Words to Literacy. 1962. Schoolmaster Publishing Company, London.

MARKS, M.R. Two kinds of regression weights that are better than betas in crossed samples. (Paper presented at the American Psychological Convention, 1966).

MICHAELIS, J.U. and TYLER, F.T. A comparison of reading ability and readability. Journal of Educational Psychology. 1951, 42, 491-498.

MILLER, G.A. Language and Communication. 1951. McGraw Hill, New York.

MILLER, G.R. and COLEMAN, E.B. A set of 36 passages calibrated for complexity. Journal of Verbal Learning and Verbal Behaviour. 1967, 6, 851-854.

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

MOSBERG, L., POTTER, T.C., and CORNELL, R.K. The relation between cloze and multiple-choice test scores as a function of relative paragraph difficulty and grade level. Research in Education. 1968. ERIC microfiche, ED 035513.

OHNMACHT, E.W., WEAVER, W.W. and KOHLER, E. Cloze and closure: a factorial study. Journal of Psychology. 1970, 74, 2, 205-217.

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

OLLER, J.W. and CONRAD, C.A. Cloze technique and ESL proficiency. Language Learning. 1971, 21, 183-195.

OSGOOD, C.E. The nature and measurement of meaning. Psychological Bulletin. 1952, 49, 197-237.

OTTO, W. and SMITH, R.J. Administering the School Reading Program. 1970, Houghton Mifflin, Boston.

POOLE, M.E. Social class differences in language predictability: written. Australian Journal of Education. 1973, 17, 3, 300-313.

POWELL, W.R. Reappraising the criteria for interpreting informal inventories. (Paper presented at the 13th Annual Conference of the International Reading Association. Boston, Massachusetts, 1968).

RANKIN, E.F. An evaluation of the cloze procedure as a technique for measuring reading comprehension. Dissertation Abstracts. 1958, 19, 4, 733-734.

RANKIN, E.F. The cloze procedure - its validity and utility. In Causey, O.S. and Eller, W. (eds), Eighth Yearbook of the National Reading Conference. 1959. National Reading Conference. Reprinted in Farr, R. (ed.) Measurement and Evaluation in Reading. 1970. Harcourt, Brace and World.

RANKIN, E.F. and CULHANE, J. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading. 1969, 13, 193-198.

RINSLAND, H.D. A Basic Vocabulary of Elementary School Children. 1945. MacMillan, New York.

RUDDELL, R.B. An investigation of the effect of the similarity of oral and written patterns of language structures on reading comprehension. Dissertation Abstracts. 1964, 24, 5207-5208A.

RUSSELL, D.H. and MERRILL, A.F. Children's librarians rate the difficulty of well known juvenile books. Elementary English. 1951, 28, 263-268.

RUSSELL, E. and THOMPSON, C. Establishing a Reading Centre: A Handbook on Remedial and Corrective Reading Instruction. 1966. North Carolina Advancement School.

SCHLESINGER, I.M. Sentence Structure and the Reading Process. 1968. Mouton Press, The Hague.

SIMONS, H.D. Reading comprehension: the need for a new perspective. Reading Research Quarterly. 1971. 6, 3, 338-363.

SPACHE, G. Contributions of allied fields to the teaching of reading. In Robinson, H.M. (ed.), Innovation and Change in Reading Instruction. 1968. Sixty-seventh yearbook of the N.S.S.E. Part 11, pp. 237-290.

SPACHE, G. Reading in the Elementary School. 1969. Allyn and Bacon, Boston.

STRANG, R.M. Reading Diagnosis and Remediation. 1968. International Reading Association, Newark, Del.

TAYLOR, W. Cloze procedure: a new tool for measuring readability. Journalism Quarterly. 1953, 30, 414-433.

TAYLOR, W. Recent developments in the use of the cloze procedure. Journalism Quarterly, 1956, 33, 42-48.

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

TREISMAN, A.M. Verbal responses and contextual constraints in language. Journal of verbal Learning and Verbal Behaviour. 1965, 4, 118-128.

VELDMAN, D.J. Fortran Programming for the Behavioural Sciences. 1967 Holt Reinhart and Winston, New York.

WARRY, R. and FITZGERALD, R.T. The new Rs in the primary school. Quarterly Review of Australian Education. 1966, 2, (whole issue).

WEAVER, W.W. Theoretical aspects of the cloze procedure. In Thurston, E.L. and Hafner, L.E., Fourteenth Yearbook of the National Reading Conference. 1965. National Reading Conference, Milwaukee.

WEINTRAUB, R. The cloze procedure. The Reading Teacher, 1968, 21, 567-571.

WESMAN, A.G. Writing the test item. In Thorndike, R.L. Educational Measurement. Second edition, 1971. American Council on Education. pp 81-129.

WILLIAMSON, H.J. AND BALL, I.L. Readability Levels of Children's Literature. 1973. Educational Resources, Melbourne.

APPENDICES

APPENDIX A

SUGGESTIONS FOR THE WRITING OF MULTIPLE CHOICE TEST ITEMS

(Wesman, 1971)

General

1. The item writer must have a thorough mastery of the subject matter being tested. Not only must he be acquainted with the facts and principles of the field he must be fully aware of their implications.

2. The writer who prepares items for use in tests of educational achievement must possess a rational and well-developed set of educational values (aims or objectives) that so permeate his thinking that he tends continually to seek these values in all his educational efforts.

3. The item writer must understand psychologically and educationally the individuals for whom the test is intended.

4. The item writer must be a master of verbal communication.

5. The item writer must be skilled in the handling of the special techniques of item writing.

6. As item writing is not a unitary skill, the item writer must be adept at writing the appropriate types of items for the subject matter being tested.

General suggestions for writing objective items:

1. Express the item as clearly as possible.
2. Wherever possible, choose words that have precise meanings.
3. Avoid complex or awkward word arrangements.
4. Include all qualifications needed to provide a reasonable basis for response selection.

5. Avoid the inclusion of nonfunctional words.
6. Avoid unessential specificity in the stem or the responses.
7. Be as accurate as possible in all parts of an item.
8. Adapt the level of difficulty of the item to the group and purpose for which it is intended.
9. Avoid irrelevant clues to the correct response.
10. Avoid stereotyped phraseology in the stem or the correct response.
11. Avoid irrelevant sources of difficulty.
12. Expose items to expert editorial scrutiny.

Specific to multiple-choice items.

1. Use either a direct question or an incomplete statement as the stem.
2. In general, include in the stem any words that otherwise must be repeated in each response.
3. Avoid negatively expressed stems if possible.
4. Provide a response that competent critics can agree on as best.
5. Make all the responses appropriate to the item stem.
6. Make all distractors plausible and attractive to examinees who lack the information or ability tested by the item.
7. Avoid highly technical distractors.
8. Avoid responses which overlap or include each other.
9. Use 'none of these' as a response only in items to which an absolutely correct answer can be given.
10. Arrange the responses in logical order, if one exists, but avoid a consistent preference for any particular response position.
11. If the item deals with the definition of a term, it is usually preferable to include the term to be defined in the stem.
12. Do not present a series of true-false statements as a multiple-choice item.

APPENDIX B

LIST OF 'KEY', 'BASIC', 'INSTANT' AND 'SIGHT' WORDS

A
about
aeroplane
after
again
against
all
almost
also
always
am
an
and
another
any
are
around
as
ask
at
auntie
away

baby
back
bad
ball
be
because
bed

been
before
being
best
better
between
big
bird
birthday
black
blue
book
boat
both
bought
box
boy
bring
brother
brought
but
buy
by

call
came
camp
can
car
cat

children
Christmas
city
clean
close
colour
come
could
course
cowboy
cut

daddy
day
dear
did
dinner
didn't
do
does
dog
doll
don't
door
down
dress
during

each
eat

end
enough
even
every

fact
far
fast
father
fell
few
field
find
fine
fire
first
fish
five
flower
fly
for
found
four
friend
from

game
garden
gave
general

get
girl
give
glad
go
going
good
got
grandma
great
green

had
hand
has
have
he
head
help
her
here
high
him
himself
his
home
hope
horse
house
how
however

I
if
in
into
is
its

jump
just

keep
kind
know

large
last
leave
left
less
let
letter
life
like
little
live
long
look
lot

made
make

man
many
may
me
men
might
more
morning
most
mother
Mr
Mrs.
much
mummy
must
my

name
near
never
new
next
night
nice
no
not
nothing
now
number

of

off
old
on
once
one
only
open
or
other
our
out
over
own

paper
part
party
people
pick
picture
place
play
please
present
pretty
public
put

rabbit
ran
read

red
right
room
round
run

said
same
sat
saw
say
school
see
seem
set
shall
she
ship
shop
should
side
since
sing
sister
sit
sleep
small
snow
so
some
something
soon

stand
start
state
still
stop
story
such
summer
sure
system

take
tea
teacher
television
tell
than
that
the
their
them
then
there
these
they
thing
think
this
those
though
thought
three
through
time

to
today
told
tonight
too
took
train
two
tree

under
until
up
upon
us
use
very

walk
want
war
was
watch
water
way
we
week
well
went
were
what
when
where

which
while
white
who
why
will
wish
with
without
woman
work
would
write

year
yes
yet
yesterday
you
your

APPENDIX C

PERCENTAGE REPLACEMENT RATES FOR EACH OF THE DELETED WORDS IN THE CLARK AND JOHNSON 1972 STUDY.

from	45.4	life	67.3
ridden	25.4	lonely	5.4
or	25.4	there	61.8
taught	58.2	from	36.4
knew	45.4	Spring	5.4
holes	65.4	the	16.4
few	54.5	of	74.5
the	70.9	are	7.3
for	91.9	their	80.0
holes	52.7	a	60.0
that	83.6	for	29.1
another	50.9	own	29.1
for	30.9	at	83.6
of	67.3	to	61.8
each	23.6	be	74.5
Doug	20.0	he	76.4
rough	7.3	better	67.3
he	71.8	first	12.7
small	15.4	good	61.8
tree	60.0	with	41.8
a	27.5	forehead	18.2
section	3.6	him	70.9
enough	32.7	but	56.4
partially	7.3	a	45.4
taught	10.9	between	25.4

APPENDIX D

SAMPLE OF CLOZE TEST USED FOR PILOT STUDY

CLOZE TEST B (Pattern 5)

DOUG OF AUSTRALIA

From the very first 1 _____ when he had	1 _____
ridden out to 2 _____ range with his father	2 _____
or one 3 _____ the stockmen, Doug had been	3 _____
taught 4 _____ of surviving in the bush.	4 _____
He 5 _____ the locations of half a dozen	5 _____
6 _____ holes in the nearby foothills.	6 _____
After 7 _____ a few more twigs on the	7 _____
8 _____ to insure the calf's safety, he	8 _____
9 _____ off to look for a drink.	9 _____
10 _____ few weeks ago the holes had	10 _____
11 _____ a little water. Doug discovered	11 _____
that 12 _____ they were quite dry. One	12 _____
after 13 _____, he lifted rocks which	13 _____
served as 14 _____ for the holes, but	14 _____
only a 15 _____ film of dampness was left	15 _____
at 16 _____ bottom of each hole. This	16 _____
was 17 _____, but not disastrous. Doug	17 _____
rested on 18 _____ granite face of the	18 _____
rough foothill 19 _____ thought for a	19 _____
moment. Then he 20 _____ back towards his	20 _____
fire and the 21 _____ calf. He stopped	21 _____

at a young 22 _____ tree which was growing 22._____

along the 23 _____. With a knife he 23 _____

managed to 24 _____ off a section of root 24 _____

with 25 _____ hollow heart containing 25 _____

enough water to 26 _____ his thirst at 26 _____

least partially. Long 27 _____ Rex the 27 _____

aboriginal stockman, had told 28 _____ 28 _____

that this tree could be a 29 _____ saver 29 _____

to a man lost in 30 _____ lonely bush, 30 _____

and Doug had never 31 _____ it. 31 _____

There were so many things 32 _____ had 32 _____

learned from him over the 33 _____. Why, 33 _____

only last Spring Rex had 34 _____ him how 34 _____

to throw the sharpened, 35 _____ boomerang 35 _____

the aborigines used instead of 36 _____ 36 _____

gun to kill the animals that 37 _____ 37 _____

found in the bush. The aborigines 38 _____ 38 _____

their own boomerangs, whittling them out

39 _____ a special type of wood, after 39 _____

40 _____ it for suppleness and strength. 40 _____

Doug 41 _____ at his own daydreaming 41 _____

because he 42 _____ left his boomerang 42 _____

at home, anyway. 43 _____ any case he had 43 _____

to confess 44 _____ no wild animals seemed 44 _____

to be 45 _____ around to provide a meal. 45 _____

Perhaps 46 _____ could find some witchetty grubs, which 47 _____ better than nothing. It was Rex 48 _____ had first explained to Doug that 49 _____ grubs were good to eat. Doug 50 _____ see him now, with his old straw hat and his jutting forehead, and he wished the stockman were with him now.

46 _____

47 _____

48 _____

49 _____

50 _____

(Note: The tests used in the Pilot Study were in a foolscap format thus the first 31 deletions were on the front of the sheet followed by "Please Turn Over the Page and Continue".)

APPENDIX E

WRITTEN INSTRUCTIONS FOR EXPERIMENTERS ADMINISTERING THE PILOT STUDY CLOZE TESTS

The purpose of this investigation is to gather information about problems associated with the use of the cloze procedure as a measure of reading comprehension.

1. Show the material to the class teacher and discuss the purpose of the testing. If the class teacher gives approval to give the tests then:

2. Hand out a copy of the test to each member of the grade according to normal classroom seating. There are, in fact, four different forms of the test and they have been placed in the envelope in groups of four. When you hand out the tests please ensure that you hand them out in this order. This will ensure that no person will have the same test, or a test from the same material, as the person beside them.

3. Read out the instruction page. Answer any questions in terms of what is in the instructions.

4. Give the children five minutes to do the sample exercise. When they have finished, briefly go over the answers explaining why these particular words were the correct replacements.

5. Ask the children to turn over the page and commence the main task. Do not directly answer any question that will give any clue to a missing word.

6. Allow the children 25 minutes to complete the task.

7. Collect all test sheets and place them back in the envelope.

8. Hand the envelope to the member of staff who comes out to the school from the College and ask him/her to return it to me.

9. If possible I would like you to do this in the first or second week of the teaching round.

Thank you for your assistance.

M.W. Boyce.

Note: The nine student teachers who administered the tests for the Pilot Study were all volunteers, and had had a briefing session at College regarding the purpose of the testing before they took the tests out to the schools. All the class teachers who were asked co-operated.

APPENDIX F

LIST OF SOURCES FOR THE PASSAGES USED FOR THE CLOZE TESTS IN THE MAIN INVESTIGATION.

School 1

Group 1: The Musical Seal. R. Farre. In High Spirits. Education Department of Victoria, n.d., pp 66-67.

Group 2: Paul Revere and the World he lived in. E. Forbes. In Wagner G.W., and Wilcox L.A. and Persons G.L. (Eds) Readers Digest Reading Skill Builder. Readers Digest Services, 1959, pp 135-136.

Group 3: The Loaded Dog. H. Lawson. In The Victorian Reader Sixth Book. Education Department of Victoria, n.d., pp 154-155.

Group 4: How Aunt Letty Killed the Panther. (Anon.) In The Victorian Reader Fifth Book. n.d., pp 131-132.

School 2

Group 1: The Claimant. In Flowerdew, P. and Stewart, S. Reading On: Yellow Book 2. Oliver and Boyd, 1963. pp 122-123.

Group 2: Meet the Smiths. In Flowerdew, P. and Stewart, S. Reading On: Red Book 1. Oliver and Boyd, 1966. pp 80-81.

Group 3: A Question of Insight. J.B. Mosley. In New Reading Skill Builder: Part 1. Readers Digest, 1968, pp 20-21.

Group 4: The Frog Prince. The Brothers Grimm. In Huber, M.B. and Salisbury, F.S. Magic Everywhere. James Nisbet and Co., 1962, pp 8-9.

School 3

Group 1: Afghanistan: Domain of the fierce and free. James A. Michener. In Scott A.F. (Ed.) New Reading: Red Book Five. Readers Digest Educational Department. 1960, pp 30-31.

Group 2: The Puddin' Thieves. Norman Lindsay. In The Planet of the Bees and Other Stories. Endeavour Reading Program 13. Jacaranda Press, 1972, pp 8-9.

Group 3: Goings On at No. 32. Michael Bond. In Reading for Pleasure. Endeavour Reading Program 11. Jacaranda Press, 1972. pp 62-63.

Group 4: Jack the story of a pretty good donkey. F.P. Jay. In New Reading Skill Builder Part 3 Sinclair, K.M. and Sparks, N.J., Readers Digest, 1971, pp 86-87.

School 4

Group 1: The Boy who wouldn't Box. In Lamb, G.F. One Hundred Good Stories.

Group 2: Christmas Trees. In Flowerdew P. and Stewart S. Reading On: Red Book 2. Oliver and Boyd, 1971. pp 8-9.

Group 3: The Seal Family. In Schonell, F.J., Flowerdew, P., and Elliott-Cannon, A. Wide Range Interest: Book 3 Oliver and Boyd, 1971, pp 124-126.

Group 4: Rip Van Winkle. In Jack and the Stolen Apples. Royal Road Readers Book 8. Daniels J.C. and Diak H. Chatto and Windus, 1970, pp 30-32.

APPENDIX G

SAMPLE OF THE CATEGORIZATION OF THE SEVEN CLOZE PATTERNS OF ONE OF THE SIXTEEN PASSAGES.

The Musical Seal

Deletion Pattern 1	No its	No syl.	In Com	P/P C/A	Deletion Pattern 2	No its	No syl	In Com	P/P C/A
Lora's*	5/6	2			musical *	7+	3		
Aunt	3/4	1	+		Miriam	5/6	3		
the *	3/4	1	+	+	piano*	5/6	3		
no *	1/2	1	+		notice	5/6	2		
wriggle	7+	2			over	3/4	2	+	+
it *	1/2	1	+	+	or	1/2	1	+	+
and	3/4	1	+	+	listen	5/6	2		
concentration	7+	3+			and *	3/4	1	+	+
swaying	7+	2			now	3/4	1	+	
body*	3/4	2			to	1/2	1	+	+
stopped *	7+	2			she*	3/4	1	+	+
minutes*	7+	2			still	5/6	1	+	
to	1/2	1	+	+	my*	1/2	1	+	+
described	7+	3			as *	1/2	1	+	
me *	1/2	1	+	+	a*	1/2	1	+	+
of *	1/2	1	+	+	songs	5/6	1		
through *	7+	1	+	+	the	3/4	1	+	+
would*	5/6	1	+		do*	1/2	1	+	
day *	3/4	1	+		For	3/4	1	+	+
a *	1/2	1	+	+	time	3/4	1	+	
wild	3/4	1			raspberries	7+	3		
animal	5/6	3			within	5/6	2		+
or *	1/2	1	+	+	two*	3/4	1	+	
of *	1/2	1	+	+	Harlech	7+	2		
a *	1/2	1	+	+	loud*	3/4	1		

Deletion Pattern 1	No its	No syl	In Com	P/P C/A
I *	1/2	1	+	+
she *	3/4	1	+	+
perhaps	7+	2		
Their	3/4	1	+	+
mewing	5/6	2		
rises	5/6	2		
treble	5/6	2		
I*	1/2	1	+	+
reedy	5/6	2		
had	3/4	1	+	
on	1/2	1	+	+
the*	3/4	1	+	+
played*	5/6	2		
slow*	3/4	1		
and*	3/4	1	+	+
to*	1/2	1	+	+
wail	3/4	1		
or*	1/2	1	+	+
annoyed	7+	3+		
grunt*	5/6	1		
flippers	7+	2		
within	5/6	2		+
get	3/4	1	+	
Danny	5/6	2		
beginning	7+	3+		

Deletion Pattern 2	No its	No syl	In Com	P/P C/A
saw*	3/4	1	+	
broke	5/6	2		
the*	3/4	1	+	+
repertoire	7+	3		
hisses	5/6	2		
from	3/4	1		
The	3/4	1	+	+
still	5/6	1		
efforts	7+	2		
the	3/4	1	+	+
her*	3/4	1	+	+
practice	7+	2		
a*	1/2	1	+	+
pace	3/4	1		
descending	7+	3+		
follow	5/6	2		
A*	1/2	1	+	+
a*	1/2	1	+	+
her*	3/4	1	+	+
and*	3/4	1	+	+
a*	1/2	1	+	+
a*	1/2	1	+	+
through	7+	1	+	+
Boy*	3/4	1	+	
to*	1/2	1	+	+

The Musical Seal

Deletion Pattern 3	No lts	No syl	In Com	P/P C/A	Deletion Pattern 4	No lts	No syl	In Com	P/P C/A
talent	5/6	2			came	3/4	1	+	
or	1/2	1	+	+	I*	1/2	1	+	+
the	3/4	1	+	+	other *	5/6	2	+	
Not *	3/4	1	+	+	so	1/2	1	+	
to	1/2	1	+	+	the*	3/4	1	+	+
more	3/4	1	+		inconveniently	7+	3+		
with *	3/4	1	+	+	an*	1/2	1	+	+
joy	3/4	1			which *	5/6	1	+	+
and *	3/4	1	+	+	then *	3/4	1	+	
the *	3/4	1	+	+	music	5/6	2		
would*	5/6	1	+		sit	3/4	1	+	
under*	5/6	2	+	+	its	3/4	1	+	+
singing	7+	2			however	7+	3		
humiliating	7+	3+			A	1/2	1	+	+
mouth *	5/6	1			organ *	5/6	2		
for *	3/4	1	+	+	a*	1/2	1	+	+
book	3/4	1	+		I*	1/2	1	+	+
a *	1/2	1	+	+	little	5/6	2	+	
the *	3/4	1	+	+	first	5/6	1	+	
when *	3/4	1	+	+	Aunt	3/4	1	+	
wild *	3/4	1	+	+	there*	5/6	1	+	
sight	5/6	1			After	5/6	1	+	+
I *	1/2	1	+	+	started*	7+	2		
To *	1/2	1	+	+	my*	1/2	1	+	+
groan	5/6	1			beside	5/6	2		+
Lora *	3/4	2			and*	3/4	1	+	+
into *	3/4	2	+	+	a*	1/2	1	+	+
largest*	7+	2			vocal	5/6	2		
includes	7+	3			grunts	5/6	1		
and *	3/4	1	+	+	a	1/2	1	+	+

Deletion Pattern 3

	No lts	No syl	In Com	P/P C/A
a *	1/2	1	+	+
roar	3/4	1		
took*	3/4	1	+	
were*	3/4	1	+	
idea	3/4	2		
own	3/4	1	+	
sessions	7+	2		
simple	5/6	2		
with	3/4	1	+	+
notes	5/6	1		
the*	3/4	1	+	+
sudden	5/6	2		
piece	5/6	1		
for	3/4	1	+	+
beat	3/4	1		
habit	5/6	2		
week	3/4	1		
Baa*	3/4	1		
without*	7+	2	+	+
learn*	5/6	1		

Deletion Pattern 4

	No lts	No syl	In Com	P/P C/A
deep	3/4	1		
turned	5/6	2		
no*	1/2	1	+	
soon	3/4	1	+	
of*	1/2	1	+	+
to*	1/2	1	+	+
that*	3/4	1	+	+
tune	3/4	1		
bars	3/4	1		
she	3/4	1	+	+
music	5/6	2		
high	3/4	1	+	
played*	5/6	2		
she*	3/4	1	+	+
about	5/6	2	+	
of*	1/2	1	+	+
she*	3/4	1	+	+
Baa*	3/4	1		
a*	1/2	1	+	+
where*	5/6	1	+	

The Musical Seal

Deletion Pattern 5	No lts	No syl	In Com	P/P C/A
out	3/4	1	+	
struck	5/6	1		
animals	7+	3+		
Lora	3/4	2		
instrument	7+	3+		
the*	3/4	1	+	+
expression	7+	3+		
was*	3/4	1	+	
with*	3/4	1	+	+
when*	3/4	1	+	+
quietly	7+	3+		
spell*	5/6	1		
can	3/4	1	+	
relation	7+	3+		
and*	3/4	1	+	+
birthday	7+	2		
decided	7+	3+		
singing	7+	2		
session	7+	2		
was*	3/4	1	+	
was*	3/4	1	+	
a*	1/2	1	+	+
off	3/4	1	+	+
annoyance	7+	3+		
me*	1/2	1	+	+
continued	7+	3+		
roar	3/4	1		
range	5/6	1		
snorts	5/6	1		

Deletion Pattern 6	No lts	No syl	In Com	P/P C/A
early	5/6	2		
up	1/2	1	+	+
would*	5/6	1	+	
she*	3/4	1	+	+
lean	3/4	1		
player's	7+	2		
of*	1/2	1	+	+
quite	5/6	1		
her*	3/4	1	+	+
the*	3/4	1	+	+
for*	3/4	1	+	+
Her*	3/4	1	+	+
only	3/4	2	+	
had*	3/4	1	+	
a*	1/2	1	+	+
present*	7+	2		
that*	3/4	1	+	+
practice	7+	2		
I*	1/2	1	+	+
out*	3/4	1	+	
not*	3/4	1	+	
preliminary	7+	3+		
on	1/2	1	+	+
I*	1/2	1	+	+
Looking*	7+	2		
singing*	7+	2		
seals*	5/6	1		
among	5/6	2		
barks	5/6	1		

Deletion Pattern 5	No lts	No syl	In Com	P/P C/A	Deletion Pattern 6	No lts	No syl	In Com	P/P C/A
Wail	3/4	1			which*	5/6	1	+	+
bass*	3/4	1			to*	1/2	1	+	+
to	1/2	1	+	+	a*	1/2	1	+	+
notice	5/6	2			but	3/4	1	+	+
outclassed	7+	3+			then*	3/4	1	+	
letting*	5/6	2			her	3/4	1	+	+
my	1/2	1	+	+	accompaniment	7+	3+		
followed	7+	3+			when*	3/4	1	+	
at*	1/2	1	+	+	a*	1/2	1	+	+
of	1/2	1	+	+	steadily	7+	3+		
made*	3/4	1	+		valiant	7+	3+		
in	1/2	1	+	+	a*	1/2	1	+	+
or	1/2	1	+	+	low*	3/4	1		
too	3/4	1	+		quickly*	7+	2		
would*	5/6	1	+		start	5/6	1	+	
with*	3/4	1	+	+	her*	3/4	1	+	+
hers*	3/4	1	+	+	when*	3/4	1	+	
was*	3/4	1	+		able*	3/4	1		
Black*	5/6	1	+		sheep*	5/6	1		
break	5/6	1			and*	3/4	1	+	+
my	1/2	1	+	+	Caravan	7+	3+		

The Musical Seal

Deletion Pattern 7	No lts	No syl	In Com	P/P C/A
whenever	7+	3+		
on*	1/2	1	+	+
take*	3/4	1	+	
would*	5/6	1	+	
against	7+	2	+	+
legs	3/4	1		
intense	7+	2		
flattering	7+	3+		
whole	5/6	1		
music*	5/6	2		
several	7+	3+		
reactions	7+	3+		
be*	1/2	1	+	
sent*	3/4	1		
book*	3/4	1	+	
Thumbing	7+	2		
I*	1/2	1	+	+
each*	3/4	1	+	
chose	5/6	1		
picking*	7+	2		
an*	1/2	1	+	+
scale	5/6	1		
Men	3/4	1	+	
heard*	5/6	1		
down*	3/4	1	+	
Whereupon	7+	3+		+
have*	3/4	1	+	
mammals	7+	2		
peculiar	7+	3+		

Deletion Pattern 7	No lts	No syl	In Com	P/P C/A
often*	5/6	2		
a*	1/2	1	+	+
hiss	3/4	1		
my	1/2	1	+	+
I*	1/2	1	+	+
sing*	3/4	1	+	
During	5/6	2		
I*	1/2	1	+	+
fairly	5/6	2		
ascending	7+	3+		
efforts	7+	2		
timeless	7+	2		
note*	3/4	1		
plainly	7+	2		
to	1/2	1	+	+
fore	3/4	1		
angry	5/6	2		
to*	1/2	1	+	+
and*	3/4	1	+	+
was*	3/4	1	+	
has	3/4	1	+	

APPENDIX H

Samples of Cloze Tests Used

1. Goings On at Number Thirty-two (Paddington Bear)
   Deletion pattern 6.

2. The Puddin' Thieves
   Deletion pattern 3.

3. Christmas Trees
   Deletion pattern 2.

4. The Boy Who Wouldn't Box
   Deletion pattern 2.

5. The Claimant
   Deletion pattern 1.

It was when he reached the top rung and peered over the edge into the loft that Paddington's worst suspicions were realised. For there, tip-toeing across 1 rafters with a torch in one 2 1 and what looked like a long . 3 . in the other, was a man 4 a trilby hat and blue overalls.

5 his breath Paddington considered the matter 6 several seconds before coming to a 7 . As quietly as possible, he stretched 8 paw into the darkness until he 9 the edge of the trap door 10 then he flung it back into 11 and pushed the bolt home as 12 as he could before scrambling down 13 ladder on to the landing and 14.

All at once there was a . 15 .. in the roof as someone started 16 shout, and several bumps followed by 17 sound of banging on the other 18 of the trap door. By that 19 Paddington was much too far away 20 hear what was going on. The 21 of the Browns' front door had 22 itself to the general hubbub and 23 was halfway down Windsor Gardens, hurrying 24 the pavement with a very determined 25 on his face indeed. All in 26, he decided that bad though his 27 . had been, things had been even 28 since he'd woken up and it 29 definitely time to call for help.

30 rounding several corners Paddington at last 31 . the place he had been looking 32. It was a large, old-fashioned . 33 building which stood slightly apart from 34 rest on a corner site. Most 35 the windows had bars across them 36 at the top of some steps 37 up to the entrance there was 38 blue lamp with the word "POLICE" 39 across it in white letters.

Paddington 40 . up the steps and then paused 41 the entrance. Leading from the hall . 42 were a number of doors and 43 was difficult to decide which was 44 best one. In the end he 45 . on a large brown door on 46 right. It looked more important than 47 of the others and Paddington was 48 firm believer in going to the 49 whenever he had an emergency.

After 50 several times he waited with his ear against the keyhole until he heard a gruff voice

"Bravely sung," exclaimed Bill, grasping Bunyip Bluegum by the hand, and they proceeded with expressions of the greatest courage and determination.

As a 1 for this renewed activity, they got 2 useful information from a Rooster who 3 standing at his front gate looking 4 and down the road, and wishing 5 heaven that somebody would come along 6 him to talk to. They got, 7 fact, a good deal more information 8 they asked for, for the Rooster 9 one of those fine upstanding, bumptious 10 who love to talk all day, 11 the heartiest manner, to total strangers 12 their wives do the washing.

"Singed 13 ," he exclaimed, when they had put 14 usual question to him. "Now, what 15 extraordinary thing that you should come 16 and ask me that question. What 17 astounding and incredible thing that you 18 actually use the word 'singed' in 19 with the word 'possum'. Though mind 20 , the word I had in my 21 was not 'singed', but 'burning'. And 22 'possum' but 'feathers'. Now, I'll tell 23 why. Only this morning, as I 24 standing here, I said

to myself ' 25 been burning feathers'. I called out 26 once to the wife — fine woman, 27 wife, you'll meet her presently — 'Have 28 been burning feathers?' 'No', says she. " 29 ", said I, 'if you haven't been 30 feathers, somebody else has.' At the 31 moment that I'm repeating the word ' 32 ' and 'burning' you come along and 33 the words 'singed' and 'possum'. Instantly 34 call to mind that at the 35 moment that I smelt something burning, 36 saw a possum passing this very 37 , though whether he happened to be 38 or not I didn't inquire."

"Which 39 did he go?" inquired Bill excitedly.

" 40 , let me see," said the Rooster. "41 went down the road, turned to 42 right, gave a jump and a 43 , and set off in the direction 44 Watkin Wombat's summer residence."

"The very 45 we're after," shouted Bill, and bolted 46 down the road, followed by the 47 , without taking any notice of the 48 request to wait a minute and 49 introduced to the wife.

"His wife 50 be all right," said Bill as they ran, "but what I say is, blow meetin' a bloomin' old Rooster's wife when you haven't got a year to waste listenin' to a bloomin' old Rooster."

They followed the Rooster's directions with the utmost rapidity, and came to a large hollow tree with a door in the side and a notice-board nailed up which said, "Watkin Wombat, Esq., Summer Residence".

CHRISTMAS TREES

No one really knows how or when the custom of the Christmas tree began, but there are many legends concerning it, and though they cannot all be true, some of them are strange and rather beautiful.

One (1) Eve in the eighth century, it (2) said, a crowd of people gathered (3) a German forest. Before them stood (4) missionary Saint Winfred who had converted (5) to Christianity. He spoke of the (6) heathen worship they had now forsaken (7) he pointed to a great oak (8) beneath which so often they had (9) human sacrifice.

"Let us hew it (10)," he said, "to show that you (11) cast away the old faith and (12) the faith of Christ." So saying, (13) began to chop down the oak, (14) soon with a mighty creak of (15) trunk and a shuddering of its (16), it crashed to earth, destroying as (17) fell every tree and plant that (18) touched. But no—not every one—(19), tall and slim and straight, stood (20) young fir tree which had miraculously (21). This tree therefore came to be (22) as a symbol of the Christ (23), the holy Christmas tree.

A story (24) France is quite different. It tells (25) the people of a certain village (26) one day that a huge fir (27) had suddenly appeared. On its branches (28) burning candles, some standing upright and (29) hanging upside down. At the top (30) the tree was a vision of (31) child, with a halo glowing about (32) head. The villagers were awed and (33), and sought an explanation from the (34) who declared that the candles represented (35) people and bad ones, and that (36) child was Christ Himself. So, from (37) time, the fir tree was used (38) Christmas to represent the Christ child.

(39) legend concerns a forester and his (40) on a stormy Christmas Eve. They (41) the door against the bitter wind, (42) gathered contentedly round the blazing fire (43) the hearth. Soon there came a (44) at the door and the father (45) it, to find a little child (46) by cold and hunger.

"Come in," (47) the father, and the family hastened (48) welcome the child. The mother fed (49), the children made room for him (50) the fire, and one of them gave him his own bed for the night. So the evening and the night passed, and the family were awakened in the morning by the singing of angel voices. In surprise they looked at their little guest and they saw that a heavenly light shone around him. Then they knew that he was the Christ

The boy who wouldn't box

"LEAD with your left, son. Lead with your left. You'l never be a real boxer if you can't do that."

Stanley lowered his gloved hands.

"But I don't want to be a boxer, 2 .. I want to be a footballer."

3 Matthews shook his head irritably.

"Nonsense, 4 , nonsense. Football won't get you anywhere. 5 footballer's finished before he's thirty. And 6 never has the time to learn 7 trade either. Now, boxing's different. You 8 be a boxer *and* run a 9 - . Look at me!"

It was true. 10 Matthews was not only a well- 11 boxer who had taken part in 12 three hundred and fifty fights, but 13 was also the owner of a 14 hairdressing business. He had long been 15 as the Fighting Barber of Hanley.

16 this time he had retired from 17 ring, but he was eager that 18 thirteen-year-old son, Stanley, should 19 to make the name of Matthews 20 important one in boxing circles. Each 21 he trained the boy to keep 22, and he encouraged him to put 23 the gloves at every opportunity. It 24 him a good deal that Stanley 25 persist in wasting his time kicking 26 football about whenever he had a . 27 moment.

Mr Matthews had no objection 28 football as a game, but he 29 not want to see his son 30 his feet when he should be 31 his fists. He made this clear 32 Stanley with increasing vigour, getting more 33 more worked up about it.

Stanley, 34 his part, grew more and more 35 . He greatly admired his father, a 36 figure in the world of sport, 37 his own interest in football was 38 passion that could not be deflected. 39 , on the other hand, made no 40 to him.

In the end worry 41 frustration made the boy literally sick. 42 mother put him to bed, and 43 had a serious talk with her 44 .

"Now listen to me," she said. "45 has got to stop. You're making 46 boy ill. Just think a moment. 47 would have happened if *your* father 48 wanted you to be a footballer 49 of a boxer?"

"I'd never have 50 up with that!" Mr Matthews admitted.

"Well, can't you see it's the same thing with Stanley, only it's football instead of boxing with him?"

Mr Matthews thought hard, and then went up to his son's bedroom.

"Look here," he told Stanley, "if you can make yourself good enough at football to be a schoolboy international before you leave school, I promise I'll do my best to help you be a professional footballer."

This was a fair offer, for Stanley was already a player of very great promise. He had played for his school (Wellington Road School, Hanley) since he was only

THE CLAIMANT

When the ship *Bella* foundered off Rio in 1854, all aboard were drowned, including Roger Tichborne, heir to a baronetcy and a fortune in Hampshire. 1 estate passed to his brother, and 2 the story might have ended had 3 not been for Lady Tichborne, who 4 refused to believe in the death 5 her son. Grief and wishful thinking 6 her to imagine that somehow Roger 7 survived and that one day he 8 return to her; and far from 9 in the course of years, her 10 grew ever stronger and more certain. 11 alone would have done no harm, 12 the poor lady inserted injudicious advertisements 13 and wide, describing her son and 14 estates to which he was heir— 15 temptation, in fact, to any rogue 16 cared to make a claim.

The 17 thing is that no-one attempted 18 deceive her earlier, and it was 19 years before an impersonator appeared--a 20 named Arthur Orton, born and bred 21 a butcher in Wapping, England, but 22 at that time in Wagga Wagga, 23 .

Arthur Orton—coarse, illiterate, dishonest—had 24 changed his name to Johnny Paisley 25 then to Tom Castro because of 26 -stealing trouble, but it is doubtful 27 he ever really intended to set 28 up as Roger Tichborne. He saw 29 of Lady Tichborne's appeals in an 30 paper and began as a mere 31 to pretend to his friends that 32 was the missing heir. He said, 33 enough, that Castro was not his 34 name and he added mysterious remarks 35 wealthier days and about South America 36 the horrors of shipwreck.

His friends, 37 enough for a little excitement, pounced 38 these signs of evidence, enlarged and 39 them and spread them in good 40 around Wagga Wagga. Feeling something of 41 hero, Arthur Orton let them talk 42 believe; and foolishly at last he 43 them to persuade him to write 44 Lady Tichborne declaring himself. The letter 45 badly written and badly spelled, but 46 Tichborne never doubted for a moment 47 it came from her long-lost 48. She at once sent him money 49 his passage to England and wrote 50 him to call at Sydney for an old Negro servant named Bogle who had worked for the family in the past.

Arthur Orton, egged on by his friends, agreed to go to Sydney. He did not really intend to leave Australia but he took the trouble to learn a little about the man he was impersonating so that he could impress Bogle with the correct "recollections". All would depend of course on whether Bogle accepted him or denounced him.

Now Bogle was a simple old man, eager to help, eager to please, and when news of "Sir Roger's" arrival in Sydney came to him he went forward in all confidence to meet him. Arthur Orton, friendly and familiar ("Well Bogle, it's a long time since I saw you"), built his conversation on the facts he had learned, and easily impressed the unsuspicious Negro. Bogle, pleased and child-like,

APPENDIX I

Number of words in each 'easy' subdivision (predictor variables) and criterion score for each cloze test.

Test	Predictor variables 1	2	3	4	5	Criterion variable.
0101	28.000	30.000	45.000	26.000	21.000	52.000
0102	31.000	35.000	44.000	31.000	24.000	44.000
0103	32.000	36.000	48.000	29.000	21.000	52.000
0104	31.000	36.000	48.000	36.000	16.000	50.000
0105	28.000	34.000	40.000	27.000	22.000	40.000
0106	30.000	35.000	45.000	31.000	22.000	64.000
0107	25.000	30.000	43.000	23.000	12.000	46.000
0201	24.000	29.000	45.000	22.000	17.000	46.000
0202	30.000	33.000	46.000	31.000	20.000	38.000
0203	32.000	40.000	45.000	30.000	23.000	32.000
0204	25.000	31.000	43.000	24.000	18.000	58.000
0205	33.000	32.000	45.000	28.000	20.000	56.000
0206	26.000	30.000	44.000	28.000	17.000	44.000
0207	29.000	32.000	47.000	24.000	18.000	40.000
0301	26.000	30.000	45.000	25.000	19.000	42.000
0302	28.000	36.000	49.000	31.000	21.000	38.000
0303	28.000	34.000	47.000	32.000	18.000	44.000
0304	30.000	36.000	49.000	26.000	17.000	28.000
0305	35.000	39.000	48.000	38.000	24.000	36.000
0306	30.000	35.000	46.000	28.000	18.000	40.000
0307	31.000	37.000	46.000	31.000	21.000	78.000
0401	24.000	34.000	49.000	24.000	15.000	16.000
0402	38.000	41.000	49.000	37.000	26.000	50.000
0403	38.000	41.000	49.000	32.000	19.000	26.000
0404	35.000	38.000	47.000	36.000	23.000	44.000
0405	33.000	38.000	48.000	30.000	18.000	22.000
0406	31.000	42.000	50.000	33.000	17.000	68.000
0407	35.000	42.000	50.000	34.000	20.000	68.000
0501	30.000	33.000	39.000	30.000	19.000	56.000
0502	26.000	27.000	46.000	30.000	20.000	76.000
0503	33.000	32.000	50.000	32.000	22.000	50.000
0504	26.000	31.000	45.000	24.000	17.000	48.000
0505	33.000	32.000	45.000	33.000	23.000	64.000
0506	28.000	31.000	44.000	31.000	17.000	58.000
0507	30.000	36.000	47.000	28.000	18.000	54.000
0601	32.000	37.000	48.000	32.000	24.000	48.000
0602	32.000	37.000	46.000	31.000	19.000	44.000
0603	24.000	35.000	44.000	22.000	15.000	24.000
0604	30.000	34.000	47.000	32.000	23.000	44.000

0605	27.000	33.000	46.000	29.000	15.000	32.000
0606	32.000	38.000	45.000	29.000	21.000	50.000
0607	30.000	37.000	45.000	29.000	19.000	36.000
0701	30.000	36.000	48.000	34.000	17.000	32.000
0702	28.000	32.000	43.000	32.000	22.000	52.000
0703	27.000	35.000	45.000	32.000	21.000	44.000
0704	26.000	30.000	44.000	28.000	12.000	52.000
0705	26.000	30.000	47.000	28.000	18.000	52.000
0706	35.000	40.000	50.000	33.000	22.000	40.000
0707	31.000	34.000	48.000	30.000	16.000	46.000
0801	41.000	45.000	50.000	44.000	25.000	58.000
0802	40.000	41.000	49.000	42.000	29.000	58.000
0803	35.000	44.000	50.000	34.000	18.000	62.000
0804	35.000	40.000	50.000	37.000	16.000	68.000
0805	45.000	43.000	49.000	41.000	32.000	84.000
0806	39.000	40.000	50.000	37.000	24.000	56.000
0807	36.000	38.000	50.000	36.000	22.000	72.000
0901	27.000	37.000	49.000	30.000	13.000	62.000
0902	36.000	37.000	42.000	34.000	20.000	58.000
0903	26.000	35.000	46.000	32.000	9.000	46.000
0904	33.000	37.000	45.000	32.000	18.000	58.000
0905	27.000	33.000	46.000	27.000	16.000	58.000
0906	32.000	37.000	46.000	31.000	19.000	48.000
0907	26.000	30.000	43.000	28.000	20.000	56.000
1001	32.000	34.000	48.000	30.000	21.000	46.000
1002	33.000	32.000	46.000	32.000	18.000	36.000
1003	35.000	37.000	47.000	35.000	20.000	46.000
1004	32.000	34.000	45.000	32.000	17.000	36.000
1005	34.000	36.000	46.000	33.000	24.000	42.000
1006	36.000	37.000	47.000	31.000	22.000	32.000
1007	35.000	39.000	44.000	30.000	16.000	38.000
1101	29.000	35.000	47.000	30.000	22.000	62.000
1102	35.000	39.000	45.000	33.000	24.000	60.000
1103	35.000	37.000	45.000	34.000	27.000	54.000
1104	30.000	38.000	48.000	33.000	19.000	76.000
1105	28.000	35.000	46.000	32.000	15.000	56.000
1106	32.000	35.000	46.000	33.000	17.000	50.000
1107	35.000	33.000	44.000	36.000	20.000	56.000
1201	33.000	34.000	49.000	33.000	16.000	60.000
1202	35.000	40.000	48.000	33.000	19.000	42.000
1203	35.000	36.000	47.000	34.000	22.000	56.000
1204	40.000	41.000	49.000	34.000	21.000	42.000
1205	34.000	40.000	50.000	33.000	16.000	42.000
1206	36.000	40.000	47.000	33.000	21.000	38.000
1207	32.000	38.000	49.000	35.000	18.000	40.000
1301	30.000	34.000	45.000	29.000	15.000	58.000
1302	34.000	37.000	47.000	32.000	21.000	72.000
1303	28.000	31.000	46.000	30.000	15.000	72.000
1304	37.000	39.000	49.000	36.000	19.000	50.000
1305	35.000	36.000	43.000	35.000	26.000	68.000
1306	39.000	40.000	50.000	38.000	12.000	72.000

1307	33.000	36.000	45.000	36.000	22.000	66.000
1401	32.000	37.000	44.000	30.000	18.000	52.000
1402	35.000	36.000	40.000	36.000	23.000	44.000
1403	33.000	37.000	48.000	31.000	21.000	24.000
1404	29.000	32.000	44.000	31.000	18.000	40.000
1406	33.000	37.000	46.000	36.000	23.000	50.000
1501	36.000	39.000	47.000	33.000	20.000	24.000
1502	34.000	36.000	47.000	31.000	16.000	36.000
1503	31.000	40.000	47.000	30.000	19.000	56.000
1504	36.000	43.000	50.000	34.000	21.000	86.000
1505	28.000	37.000	46.000	25.000	11.000	14.000
1506	30.000	33.000	48.000	32.000	21.000	38.000
1507	32.000	36.000	47.000	35.000	15.000	42.000
1601	37.000	40.000	49.000	34.000	18.000	32.000
1602	33.000	35.000	47.000	32.000	16.000	48.000
1603	34.000	37.000	48.000	25.000	14.000	30.000
1604	34.000	35.000	46.000	29.000	18.000	44.000
1605	33.000	38.000	47.000	29.000	18.000	16.000
1606	38.000	39.000	48.000	34.000	22.000	18.000
1607	32.000	35.000	49.000	26.000	11.000	26.000

* Predictor variables

1. Words of 1-4 letters.
2. Words of 1 syllable.
3. Words of 1-2 syllables.
4. Words 'in' common word lists.
5. Articles, conjunctions, prepositions and pronouns.

APPENDIX J

SAMPLE OF THE DETAILED RESULTS FOR EACH PATTERN FOR TWO OF THE SIXTEEN PASSAGES

(a) The Musical Seal. School 1 Passage 1

Mean score 24.85 (49.70%)

Category 1: Number of letters

Number of letters/number correct.

Excerpt	1/2		3/4		5/6		7+	
1	14	12	14	7	11	4	11	3
2	12	10	19	10	12	1	7	1
3	6	4	26	16	12	4	6	2
4	14	11	17	7	16	6	3	1
5	9	3	19	12	8	4	14	1
6	10	8	20	16	9	4	11	4
7	10	8	15	10	10	4	15	1
	75	56	130	78	78	27	67	13
%age correct		74.67		60.00		34.62		19.40
%age of whole		21.43		37.14		22.29		19.14
%age of correct		32.18		44.83		15.52		7.47

Category 2: Number of syllables

	1		2		3+	
1	30	21	15	5	5	0
2	35	20	9	0	6	2
3	36	21	12	5	2	0
4	36	21	12	4	2	0
5	34	18	6	2	10	0
6	35	27	10	5	5	0
7	30	20	13	3	7	0
	236	148	77	24	37	2
%age correct		62.71		31.17		5.41
%age of whole		67.43		22.00		10.57
%age of correct		85.06		13.79		1.15

Category 3: Common 200 words

	In		Not	
1	26	19	24	7
2	31	19	19	3
3	29	20	21	6
4	36	21	14	4
5	27	16	23	4
6	31	25	19	7
7	23	17	27	6
	203	137	147	37
%age correct		67.48		25.18
%age of whole		58.00		42.00
%age of correct		78.74		21.26

(b) How Aunt Letty Killed the Panther. School 1 Passage 4

Mean score 21.00 (42.00%)

Category 1 Number of letters

Deletion Pattern	No wds	No Corr 1/2		3/4		5/6		7+	
1	7	3	17	4	16	0	10	1
2	8	6	30	18	5	1	7	0
3	8	1	30	11	7	1	5	0
4	5	4	30	16	10	2	5	0
5	8	2	25	6	7	3	10	0
6	4	2	27	22	15	9	4	1
7	7	7	28	21	12	6	3	0
	47	25	187	98	72	22	44	2
%age correct		53.19		52.40		30.55		4.54

How Aunt Letty Killed the Panther (cont'd.)

Category 2. Number of syllables

	1		2		3+	
1	34	7	15	1	1	0
2	41	25	8	0	1	0
3	41	12	8	1	1	0
4	38	19	9	3	3	0
5	38	9	10	2	2	0
6	42	28	8	6	0	0
7	42	31	8	3	0	0
	276	131	86	16	8	0
%age correct	47.76		26.67		0.00	

Category 3: In common word lists

	In		Not In	
1	24	7	26	1
2	37	24	13	1
3	32	11	18	2
4	36	20	14	2
5	30	8	20	3
6	33	25	17	9
7	34	28	16	6
	226	123	124	24
%age correct	54.43		19.36	

Category 4 Part of Speech.

	Prep/Pro Conj/Art		Other	
1	15	6	35	2
2	26	20	24	5
3	19	5	31	8
4	23	12	27	10
5	18	6	32	5
6	17	14	33	20
7	20	18	30	16
	138	81	212	66
%age correct	58.70		31.14	