Edith Cowan University

Research Online

1993

The effects of different methods of cloze test construction and their relationship with a standardised reading comprehension test

Trevor Michael Edward Forde
Edith Cowan University

Follow this and additional works at: https://ro.ecu.edu.au/theses_hons

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation

THE EFFECTS OF DIFFERENT METHODS

OF CLOZE TEST CONSTRUCTION AND

THEIR RELATIONSHIP WITH A

STANDARDISED READING COMPREHENSION TEST

BY

Trevor Michael Edward Forde, BA (Ed.)

A Thesis Submitted in Partial Fulfilment of

the Requirements for the Award of

Bachelor of Education with Honours

at the School of Education,

Edith Cowan University

Date of Submission: 19.02.93

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

Abstract

The purpose of this study is to provide guidelines to classroom teachers for the construction of valid and reliable cloze comprehension tests. Review of the related literature suggested there are many diverse and contradictary opinions as to the most appropriate method of constructing cloze tests.

The subjects were 49 Year 5 primary school children from two metropolitan schools. Three different comprehension tests were administered. The <u>Gap Reading Comprehension Test</u> and two Cloze Comprehension Tests consisting of 50 and 100 word deletions, in which the deletions began at either the fifth or sixth word of the second sentence. The Cloze Tests were graded with both the exact word and synonymic word scoring methods and then correlated with the <u>Gap Reading Comprehension Test.</u> Reliability was established by the split-half method.

Concurrent validity and reliablity were found to be greater in the 100 word deletion versions of the Cloze Tests, with increased reliability achieved when the tests were graded with the synonymic word scoring method.

The results provide useful guidelines for classroom teachers in the construction of valid and reliable measures of reading comprehension.

ii

Declaration

I certify that this thesis does not incorporate, without acknowledgement, any material previously submitted for a degree or diploma in any institution of higher education and that, to the best of my knowledge and belief, it does not contain any material previously published or written by another person exept where due reference is made in the text.

Trevor Forde

School of Education

Edith Cowan University

Churchlands W.A. 1992

Acknowledgements

List of Tables

Table of Contents

Chapter 1

Introduction

The Background to the Study

Cloze procedure was first introduced by Wilson Taylor in 1953 as a means of measuring readability of texts. It has three main uses in the classroom today: as an instructional tool, as a way of determining readability, and as a measure of comprehension. The focus of the present study is on cloze as a measure of comprehension. The Bullock Report (D.E.S. 1975, cited in Rye, 1982, p.1) described the process of using cloze as a measure of comprehension as "...the use of a piece of writing in which certain words have been deleted and the pupil has to make maximum possible use of context clues available in predicting the missing words". The comprehension ability of the student is assessed on the basis of their ability to scan the surrounding context to obtain meaning and replace the missing word.

The generally accepted standard cloze comprehension test, first proposed by Taylor (1953), contains 50 deletions with the first and last sentences left intact. The deletions begin with the fifth word of the second sentence and continue with the deletion of every fifth word throughout the text. Tests are scored by means of marking the insertion of the original word from the text as correct and all other words as incorrect.

Research has challenged this accepted standard format of cloze comprehension testing. Sciarone & Schoorl (1989) have challenged the recommended use of 50 word deletions as being inadequate to test comprehension. Bormuth (1963; 1964), Meridith & Vaughan (1978), Porter (1978), and Taylor (1954), (cited in Helfeldt, Henk, & Fotos, 1986,

1

p.216) have shown that when five different cloze forms are constructed on a passage using the every fifth word deletion method, by varying the starting point for deletions between the fifth and ninth words of the second sentence, they yield five different results. The current emphasis in reading for meaning disputes the use of the exact word method of grading cloze tests, as the meaning of the text can be retained by using appropriate synonyms, yet would be marked incorrect using the standard cloze format. Henk (1981, p.348) cited four studies, Anderson (1972), Hargis (1972), McKenna (1976), and Ruddell (1964), where no appreciable difference in the ranking of the students occurred through the use of synonymic scoring, where words which retained the semantic and syntactic structure of the text were scored as correct.

The diverse range of opinions would suggest that there is some disagreement in deciding on the most appropriate method for achieving validity and reliability in cloze test construction.

The Purpose of the Study

The purpose of the study was to provide guidelines for classroom teachers in the construction of valid and reliable cloze comprehension tests. To achieve this aim, the accepted standard 50 word deletion method and the alternative 100 word deletion method were correlated with the standardised Gap Reading Comprehension Test, varying the starting point for deletions between the fifth and sixth words of the second sentence of the passages. The study also investigated the relationship between the exact word and synonymic word methods of scoring cloze tests and their relationship with the Gap Reading Comprehension Test.

It was anticipated that the results from the study would assist teachers in their use of cloze tests in the following two areas:

2

1) to use the minimum number of word deletions to ensure valid and reliable comprehension test results;

2) to provide direction for the use of synonyms when grading cloze comprehension tests.

Statement of the Problem

Cloze testing of comprehension has been in use since 1953 yet there are still many opinions on how best to construct cloze and others would question its use at all (Shanahan, Kamil & Tobin, 1982). With so many diverse opinions on its use, two questions are posed for the classroom teacher. First of all, should I use cloze testing? Secondly, if so, which approach should I use?

The major advantage of cloze is its simplicity in construction in the practical classroom situation. It provides an alternative means for testing comprehension to the professionally developed standardised reading tests. Many of these tests require much time, effort, and money in order to ensure their reliability and validity for classroom use. It is not to be expected that classroom teachers would have these resources available for the construction of their own standardised comprehension tests and so they turn to cloze. With so many differing opinions on cloze construction methods, can teachers be assured that the cloze method they are using is a valid and reliable measure of comprehension? How can teachers ensure that the construction of their test provides a high level of validity and reliability in the classroom situation?

3

The following terms have special relevance to the study and are defined in the following way.

Reading Comprehension

Reading comprehension is defined as an interactive process in which "meaning is built in the mind of the reader on the basis of information taken from the reader, the text, and the reading context" (Lipson & Wixson, 1991, p.13). For the purpose of this study, reading comprehension will be assessed by the reader's ability to successfully complete cloze deletion exercises on the Gap Reading Comprehension Test.

Cloze Procedure

Cloze procedure is defined as a procedure in which certain words have been deleted from a passage. The reader has to make the best possible use of context clues in order to predict the missing words (The Bullock Report, 1975, cited in Rye, 1982, p.1). In this study the procedure involved presenting to the children passages containing 50 word and 100 word deletions. They were then asked to insert the word they felt should go in each blank space.

Exact Word Scoring Method

This is defined as the scoring procedure used to calculate the children's scores in which only the replacement of the exact word deleted from the passage is scored as correct.

Synonymic Word Scoring Method.

This is defined as the scoring procedure used to calculate the children's scores where the replacement of the deleted word by an appropriate synonym, which retains the meaning of the passage, is scored as correct.

Standardised Reading Tests

Standardised reading tests have been defined as tests that have been developed by experts and where the individual items have been analysed and revised until they meet given standards of quality. Directions for administering, scoring and interpreting are carefully specified to ensure objectivity, whereby a student's score is the same regardless of who is marking the test. Validity and reliability data are provided (Gay, 1981).

Validity

Validity is defined as the degree to which a test measures what it is intended to measure (Gay, 1981).

Content validity is defined as the degree to which a test measures an intended content area and is determined by expert judgement.

Concurrent validity is defined as the degree to which scores on a test are related to the scores on an already established test. This is determined by administering the two tests to the same group of children and correlating the two sets of scores (Gay, 1981).

Reliability

This is defined as the degree to which a test consistently measures what it is supposed to measure.

Split-half reliability is defined as a type of reliability based on the internal consistency of the test (Gay, 1981).

Parallel Forms

Parallel forms are defined as the construction and administration of two forms of the same test to the same group of people (Gay, 1981).

Statement of the Research Questions

1) Is the 50 word deletion method of Cloze Comprehension Testing a valid measure of reading comprehension when:

    a) The exact word scoring method is used?

    b) The synonymic word scoring method is used?

2) Is the 100 word deletion method of Cloze Comprehension Testing a valid measure of reading comprehension when:

    a) The exact word scoring method is used?

    b) The synonymic word scoring method is used?

3) Is the 50 word deletion method of Cloze Comprehension Testing a reliable measure of reading comprehension when:

    a) The exact word scoring method is used?

    b) The synonymic word scoring method is used?

4) Is the 100 word deletion method of Cloze Comprehension Testing a reliable measure of reading comprehension when:

    a) The exact word scoring method is used?

    b) The synonymic word scoring method is used?

5) To what extent do measures using the synonymic word method of cloze grading correlate with the exact word method of cloze grading?

The Significance of the Study

Many studies have been conducted into the validity of cloze as a means of testing comprehension. Research has focused on relevance of text choice (Alderson, 1980), trait structures (Bachman, 1982), factor validity (Bormuth, 1969), effect of grade level on cloze scores, (Dufflemeyer, 1983), modified deletion strategies (Henk, 1981), construct validity (Shanahan & Kamil, 1983), and sensitivity to intersentential constraint (Kibby, 1980; Rye, 1984; Shanahan, Kamil, & Tobin, 1982). These studies have provided conflicting evidence on the construction and effectiveness of using cloze procedures.

There are though, only a limited number of studies conducted into the number of deletions necessary to achieve validity and reliability for cloze comprehension tests and these yield conflicting results. Sciarone and Schoorl (1989) recommended 100 deletions as a minimum, while Bachman (1985), and Rand (1978) concluded maximum reliability had been achieved by 30 and 25 items respectively.

When Taylor (1953) first introduced cloze he posed the question, "How many blanks are enough? - a matter to be settled by experiment" (p. 148). The standard acceptance of a maximum 50 deletions does not appear to have been resolved by experiment, but rather by Taylor's (1956) claim that "a series of 50 blanks is roughly sufficient to allow the chances of mechanically selecting easy or hard words to cancel out..." (p.48).

If it is possible to obtain five different results from different cloze construction of one passage (Helfeldt et al., 1986), by starting the every fifth word deletions between the fifth and ninth words of the second sentence, the reliability of this simple measure is in question.

8

Thus the acceptance of synonyms in cloze grading, as opposed to exact word replacement, is in keeping with current reading theory in which reading is seen as a process of interaction between the reader, text and context (Lipson & Wixson, 1991). Researchers (Schell, 1988; Sternberg, 1991; Wood, 1988) have questioned the ability of cloze to test comprehension. As Henk (1981) pointed out, "Conceivably an individual could fully comprehend a passage and still score in the frustration range if enough responses, even though syntactically and semantically appropriate, were not exact matches"(p.348).

A major flaw in the previous research into cloze construction has been the use of university undergraduates or foreign language learners as subjects in the research. Of the research discussed in the review of the literature only Bormuth (1969) and Heldfldt et al. (1986) used primary school children as subjects. Even Taylor's (1953 ; 1956) initial research into the cloze procedure used journalism students at the University of Illinois and trainees at the Sampson Air Base in New York. Serious concerns can be raised about the generalisation of results from experienced readers to younger inexperienced readers still in the process of developing appropriate reading strategies.

The main advantage of cloze procedures in the classroom, is the relative ease with which a teacher may construct a comprehension test that correlates highly with related performances on standardised reading tests. With so many diverse opinions in the literature it is difficult for teachers to decide on the most appropriate method of constructing a Cloze Comprehension Test. This study attempts to redress this problem by investigating the following four issues, using as subjects Year 5 primary school students.

1) The number of deletions required to achieve reliability.

2) The effects of different starting points for deletions on the validity and reliability of Cloze Tests.

3) The number of deletions required to achieve parallel results when using different texts.

4) The comparative grading of Cloze tests with the exact word scoring method and the synonymic scoring method.

It is anticipated that the results of this study will act as a guideline for teachers in their search for the best use of their time and resources in constructing valid and reliable Cloze Comprehension Tests.

Limitations of the Study

The following factors are noted as limitations affecting the   findings reported in this study.

1) The passages were not chosen randomly but were chosen for their appropriateness to the Year 5 level. The content of the passages may not provide the same degree of interest to all the children in the study.

2) The children chosen to participate in the study were not chosen randomly. Due to the inaccessability of schools selected by stratified sampling, the two schools used in the study were chosen for their availability.

3) The results of the study can only be generalised to the year five primary school level on a passage of a similar difficulty level.

4) The words in the passage have been deleted using the fixed ratio deletion method in which every fifth word in the passage, beginning at either the fifth or sixth word of the second sentence, has been deleted. According to Helfeldt et al., (1986) this will cause limitations to the results of the study as not every word in the passage had an equal chance of deletion.

5) The results of all the subtests can only be generalised to tests using fixed ratio every fifth word deletions beginning the deletions at the fifth or sixth word of the second sentence.

A problem with all testing is that the assessment requires interpretation based on an individual's performance on a given text in a given context. The results will be influenced by the nature of the task, the context in which the task is given and the reader's prior knowledge and reading abilities. (Johnston, 1983, p.20). Schell (1988) suggested all teachers should be careful in their evaluation of comprehension tests and stated, "Maybe all we can say in some circumstances is that a reader had trouble comprehending specific material under certain conditions" (p. 13).

Plan of the Thesis

The investigation is reported according to the plan set out below.

Chapter 2

Chapter 2 provides a literature review of previous research related to this study.

Chapter 3

Chapter 3 discusses the theoretical rationale for the use of the cloze procedure.

Chapter 4

Chapter 4 describes the methods of investigation including the design of the study, the samples and instruments to be used and the data collection and analysis procedures,

Chapter 5

Chapter 5 presents an analysis of the results.

Chapter 6

Chapter 6 presents a discussion of the results, conclusions and implications for classroom teachers.

Chapter 2

Review of the Literature

Reading

Theory on how children learn to read has traditionally been based on two diverse perspectives. The first emphasises the importance of the text, commonly known as `Bottom Up' or `Outside In' processing where meaning develops from sound/symbol, to word, to sentence and finally to text where meaning is discovered. The emphasis in the `Bottom-Up' theory is on the information the reader extracts from the page. This is the information that a reader obtains through the eyes while looking at a text. Quite simply, the information available to you until you turn off the lights (Smith, 1985).

The second perspective, known as `Top Down' or `Inside Out' processing, emphasises that meaning is the starting point for reading and directs all other activities (Sloan & Whitehead, 1986). The `Top Down' theory of reading refers to the information readers already possesses through their prior knowledge and experiences. This is the knowledge they possess relating to the language itself and their experiences and understandings of the world. The reader uses his/her prior knowledge of three language systems to predict and confirm the meaning of the text (Lipson & Wixson, 1991, p.9). They are as follows:

1) Grapho-phonic information - the relationship between sounds and symbols in a text.

2) Syntactic information - an understanding of how the language functions.

3) Semantic information - the experiences and understandings   individuals have had

which they bring to the text (Latham & Sloan, 1979).

Parker (1985) argues that a proficient reader uses all three language systems in an integrated and automatic way.

Smith (1985) used the terms visual and non-visual information to describe `Bottom-Up' and `Top-Down' processing respectively. Smith (1985) argued that visual and non-visual information are not used in isolation, but in fact the relationship is a reciprocal one. He stated, " The more non-visual information you have when you read the less visual information you need and the less non-visual information you have when you read the more visual information you need" (p.14). This can be demonstrated when trying to read a foreign language or an unfamiliar text such as a medical journal. Although we can see the letters and possibly pronounce the words, unless we have prior knowledge of the foreign language or a medical background we will not obtain the meaning. In order to understand these texts we will use more visual cues, such as syllabification, to try to obtain meaning.

The most recent theory, `Interactive processing', states that reading is not either a `Bottom-Up' or a`Top-Down' process but, "...is an interactive processs in which bottom-up and top-down processes occur simultaneously and meaning results from the interaction between the reader and the text" (Lipson & Wixson, 1991, p.11). This theory suggests readers can be taught to adjust their reading strategies to select the best strategy for their purpose for reading and the demands of the text (Sloan et al., 1986, p.7). The basis for this theory is that readers construct meaning when they comprehend through the interaction of three major factors; the reader, the text, and the context. Reader factors are defined as prior knowledge, knowledge about reading, and attitude and motivation for reading. Text factors are defined as the type and organisation of the text, the linguistic properties of the text, and the structural features of the text such as headings and maps. Context factors are defined as the purpose and

14

task for reading, the general and specific settings in which the reading and/or instruction occur, and the instruction itself, both content and methodology (Lipson & Wixson, 1991).

Thus, in summary, reading can be seen as an active process in which meaning is constructed through the interaction of the reader, text and context. Readers use their background knowledge in an interactive process with print information in order to obtain meaning. The extent to which a reader uses either visual or non-visual information is dependent on the amount of background knowledge the reader brings to the text.

Comprehension

The diverse range of reading theories has implications for the way we assess comprehension. If we do not have a clear and agreeable definition on what readers do when comprehending, then confusion can reign in the construction of tests to assess children's comprehension ability. Comprehension is such a complex process that assigning a simple definition to the process is extremely difficult, if not impossible. Indeed one is justified in asking if there is a definition of comprehension independent of one's own point of view. Based on Lipson & Wixson's (1991) view, I have defined comprehension as an interactive process in which a reader obtains meaning through the interaction of the reader with the text and the reading context.

In 1980 a study was conducted by Greenlaw and Kurth, in the United States, to determine if a definition of comprehension was widely held by teachers in the elementary and secondary school levels. One hundred and forty seven teachers responded to a questionnaire ranking eight definitions of comprehension from most agreeable to least agreeable. They showed a distinct preference for Jack Holme's and Russell Stauffer's models (Singer & Ruddell, 1970, cited in Greenlaw et al., 1980, p.5) which equate reading to an intellectual

15

process similar to thinking (Top Down). The least chosen models were George Miller's (Singer & Ruddell, 1970, cited in Greenlaw et al., 1980, p.4) and Wayne Otto's (1977) models which represented reading comprehension as a series of isolated skills (Bottom Up).

Irwin (1991) refers to the isolated skills approach as the way in which comprehension has been traditionally taught. Citing Rosenshine (1980, p.2) Irwin asserts there is little research to support the theory that separable skills exist in the first place. In contrast to attempts to describe isolable subskills, Irwin presents a model of what occurs when a reader is comprehending:

> 'Comprehension can be seen as the process of using one's own prior experiences and the writer's cues to construct a set of meanings that are useful to the individual reader reading in a specific context. This process can involve understanding and selectively recalling ideas in individual sentences (microprocesses), inferring relationships between clauses and sentences (integrative processes), organising ideas around summarising ideas (macroprocesses), and making inferences not necessarily intended by the author (elaborative processes). These processes work together (interactive hypothesis) and can be controlled by the reader as required by the reader's goals (metacognitive processes) and the total situation in which comprehension is occurring (situational context). When a reader consciously selects a process for a specific purpose, that process can be called a reading strategy.'(p.9)

This transactional definition of comprehension is in keeping with the interactive theory of the reading process (Lipson & Wixson, 1991) in which there is an interaction between the writer's cues and a reader's prior knowledge in a specific context.

Cloze Procedures

In many traditional tests of comprehension, including diagnostic tests and standardised tests, comprehension is measured by cloze testing. Cloze was first introduced as the Ebbinghaus Completion Method in 1897 and involved missing word techniques and sentence completion (De santi, 1985). The standard cloze procedure used in many classrooms today was introduced by Wilson Taylor in 1953. Since then cloze has been used to determine the readability of materials, as an instructional tool, and as a measure of reading comprehension.

The cloze procedure involves deleting every Nth word from a passage and replacing the word by a standard length blank. This process is known as fixed-ratio deletion cloze as opposed to rational deletion cloze in which specific words from the passage are deleted for a specific purpose. In both methods the first and last sentences in the passage are left intact to allow suitable context clues before and after the deletions. The cloze tests are graded either by the exact word method, in which the exact word deleted from the passage has to be restored in order to be scored correct, and the synonymic word method, which requires a contextually appropriate synonym to be inserted in order to be scored correct.

Research conducted by Fletcher (1959), Jenkinson (1957), and Smith & Zinc (1977), (cited in Heldfeldt, Henk, & Fotos, 1986, p.216) has shown that cloze, as a measure of comprehension, correlates highly with results achieved on standardised reading tests. The major advantage cloze has in this area is the relative simplicity of construction, compared to the more time consuming construction of many other (standardised) reading tests. The individual items in standardised tests are tested and re-tested on large numbers of subjects until each item reaches a certain standard of quality. This requires a great deal of time and money. From a

practical point of view cloze tests are also easier to mark, if the exact word grading method is employed, compared to some of the more complicated standardised reading tests.

There is some concern   (Schell, 1988; Sternberg, 1991; and Wood, 1988) about the ability of cloze to effectively test what the teachers are trying to teach. They argued that current reading tests provide an incomplete and distorted view of student performance and advise caution in the use of test scores.

This is a view supported by Irwin (1991, p.194) who stated that while comprehension tests provide an indication of how well students comprehend compared to their peers, they do not tell the teacher what instruction to provide. She also pointed out serious limitations of standardised testing. Time restraints ensure the use of short passages, different types of passages which are usually mixed together, and prior knowledge is not generally assessed or controlled.

A second major use of the cloze procedure in the classroom is in determining readability. This is concerned with matching the reader with a text at an appropriate reading level. Traditionally this has been achieved through the use of readability formulae. These are usually based ᴏₙ the average sentence length and the average word complexity (Zakaluk and Samuels, 1988, p.36). These formulae have been the cause of some concern   due to the variability of results from different formulae (Bormuth, 1966; Gray and Leary, 1935; McLaughlin, 1968; Taylor, 1953; cited in Gilliland, 1972). Lipson and Wixson (1991) stated that differences of two or more grade levels are not uncommon (p.404).

The relationship of cloze procedures to readability formulae has been put forward by Klare (1984, cited in Lipson et al., 1991 p.408) who stated, "Formulas predict readability; cloze procedure and other similar comprehension methods measure readability".

18

Cloze Construction

The standard cloze test contains 50 word deletions from a text in which the first and last sentences remain intact in order to provide context clues before the first deletion and after the last deletion. The deletions begin with the fifth word of the second sentence and continue through the text with every other fifth word deleted. Thus a passage of approximately 300 words is required to allow for 50 times 5 words plus the first and last sentences. This form of deletion is known as fixed ratio deletion as the deletions are made according to a predetermined ratio of every nth word.

Another form of deletion, in which specific word types, such as nouns, prepositions and verbs, are deleted for specific teaching purposes, is known as variable ratio or rational deletion, as there is no predetermined ratio and specific words are deleted for a specific purpose (Soudek et al., 1983, p.336).

A problem with the fixed-ratio deletion method is that researchers have shown that five different scores can be obtained on the same passage by varying the starting point for deletions between the fifth and ninth words (Bormuth 1964; Meridith and Vaughan 1978; and Porter 1978; cited in Heldfelt et al., 1986, p.216). These different forms did not yield equivalent results due to the different types of words deleted in each version of the passage which either increased or decreased the difficulty of the passage.

Random Deletions

Alderson (1980) suggested differences in cloze test scores may not be due to differences in deletion frequency, but to differences in the particular words deleted. He argued, random deletion ignores the syntactical-semantic relationship in a text and the inconsistency in results will depend on what proportion of syntactic and textual function words are deleted. Bachman (1982) concluded, that cloze measures using rational deletions could be used to measure textual relationships beyond clause boundaries and measure higher order skills such as cohesion and coherence. In a later study (1985) Bachman found random deletion tests were comparable in reliability and validity to fixed ratio tests (p.550).

Helfeldt et al., (1986) supported Bachman's findings in their study of different cloze tests with 6th grade students. Using random deletions, by assigning consecutive integers to words between the first and last sentence in conjunction with initial letter cues, they obtained higher reliability estimates than when standard cloze tests were used. They asserted that the total random cloze more closely resembled reading, in the way points of uncertainty were more likely to occur at random(p.221). That is to say that a reader will not experience difficulty in a text at every nth word but this is more likely to occur at random.

Henk (1981) identified two major reasons why researchers were reluctant to implement total random deletions. They are more difficult to implement and the blank spaces may be too close in proximity to allow enough context for the reader to attain meaning.

The arguments for random deletions contradict Taylor (1953, p. 419) when he stated, "A random deletion which ignores the differences between specific words appears to be not only defensible but rationally inescapable when cloze procedure is used for contrasting

readabilities....If enough words are struck out at random the blanks will come to represent proportionally all kinds of words to the extent they occur".

Intersentential Constraint

The ability of cloze to be sensitive to intersentential constraint, using information across sentence boundaries, has also been called into question. Kibby (1980), Shanahan and Kamil (1983), and Shanahan, Kamil, & Tobin (1982) have demonstrated that cloze is insensitive to integrating information across sentence boundaries. This research, they assert, demonstrated that cloze has low construct validity as a means of testing reading comprehension.

Kibby (1980) tested mature readers on two paragraphs of varying difficulty presented in three formats, regular cloze, same paragraph with sentences in scrambled order and sentences read in isolation. No significant differences were found between student's performance on the regular and scrambled cloze, but in the isolated condition their performance was 10 to 15% lower. In questioning the construct validity of cloze, Kibby concluded that his study indicated cloze is largely a measure of sentence comprehension (p.310) while hypothesising that cloze might also measure only literal information (p.309).

Shanahan, Kamil, and Tobin (1982) used three different cloze formats; standard cloze, same passages with sentences scrambled, and sentences from original texts in non-supportive text, in order to investigate the ability of cloze to measure the use of information across sentence boundaries. The passages were either lengthened or shortened as necessary to ensure all sentence lengths were equal to multiples of five. These passages were then administered to 125 university undergraduates. Based on the results, which failed to distinguish between groups which had been presented with the sequential passages and those

which had scrambled passages, Shanahan et al., concluded cloze was not a good measure of intersentential comprehension.

The validity of these findings were questioned by Henk (1982) on the basis of two major issues. First of all, the manner in which the text was added to or words omitted in order to conform to the multiple of five was brought into question. According to Henk (p.589) "Such a disruption would almost certainly influence the unique cognitive-linguistic interaction operating between encoders and decoders in any communication setting." Cziko (1983, cited in Rye, 1985, p.102) also raised the possibility of sampling bias concerning the possible destruction of semantic and syntactic integrity when the passages were revised. He also added that the original passages had low intersentential constraint in the first place. Second, Henk (1982) questioned the generalisation of results obtained through testing university undergraduates, to younger, inexperienced readers (p.592).

In response to the construct validity of cloze being called into question, and the far reaching implications for teachers and researchers if it were true, Rye (1984) examined the sensitivity of cloze to intersentential constraint using 70 further education college students. One group was presented with a standardised cloze test; the other with the same material with the sentence order randomised. Rye's study differed from previous research in that the standard cloze group were given as much time as they required to complete the test. The group receiving the randomised cloze test were presented with three to five sentences printed on each page. Once they had turned the page they were not allowed to go back and alter previous material. This was to prevent this group from restructuring the passage.

Rye's results showed the group which completed the standard cloze test yielded superior results, due, he stated, to the availability of intersentential constraint (p.120). He also suggested, the results probably underestimated the true difference as it was impossible to

remove all intersentential constraint from cloze passages (p.120).

Grading Cloze Tests

The most appropriate method with which to grade a cloze passage has also stimulated some debate. There are two methods for grading cloze tests: The standard cloze exact word method, which requires the reader to replace the exact word from the text, and the synonymic word method which allows a contextually appropriate synonym to be replaced in the passage.

Most studies, Alexander (1968), Rankin (1958), Smith and Zinc (1977), and Wiechelman (1971, cited in Henk & Selders, 1984, p.282), support the use of the exact word method due to its high correlation with standardised reading tests. It is argued the results will remain objective across a variety of tests regardless of the tester. Taylor (1956, p.48) stated there was no advantage to be gained by going to the trouble of judging and scoring synonyms.

Henk (1981), while ackowledging that verbatim scoring eliminates subjectivity, asserted that using the exact word method would present the possibility for a student to score in the frustration range, even though the responses were semantically and syntactically appropriate, if enough responses had not replaced the original word from the text.

Scirone and Schoorl (1989, p.433) supported this view by suggesting that because there was no reason to expect the reader to have exactly the same sense of style as the author, then there was no reason to expect the reader to score highly in terms of exact responses.

23

It is the subjectivity of grading synonyms that causes concern. Henk et al., (1984, p.284), showed an individual's synonymic cloze score could vary by as much as 25 percentage points. Despite research to the contrary, Anderson (1972), Bormuth (1965), Miller and Coleman (1967, cited in Henk et al., 1984, p.286) and Henk (1981) recommend that responses that preserve meaning should be counted. This should be done through increased training and guidelines for grading synonyms.

In studies conducted by Anderson (1972), Hargis (1972), McKenna (1976), and Ruddell (1964, cited in Henk, 1981, p.348) no appreciable differences in rank were obtained through synonymic scoring. Henk himself found increased performances were obtained by synonymic scoring, if used in conjunction with every fifth word deletions and first letter cued blanks. Henk and Selders (1984) showed synonymic scoring by itself to be highly variable and too dependent on who scored the test. Despite their findings providing little evidence to suggest there is any advantage in doing so, they recommended responses which preserve meaning should be counted.

De Santi and O'Sullivan (1985) agreed with Henk et al., proposing the need for subjectivity in evaluating a reader's response. The emphasis on reading for meaning would seem to support the use of synonymic scoring.

Number of Deletions

Research into the number of deletions required to adequately test comprehension has been relatively limited. Bachman (1985) found cloze tests did not need to be as long or deletions as frequent as recommended in the literature in order to be reliable. His tests included only 30 deletions with a ratio of 1 deletion every 11 words. These findings were consistent with

24

Rand (1978) who found that maximum reliability had almost been achieved by 25 deletions, though the best results were achieved through testing with 50 items. He recommended the most efficient use of everyone's resources would be through the use of 25 deletions in combination with the use of the synonymic word method of grading.

Bormuth (1964) (cited in Rand, 1978, p.62) found tests with less than 50 deletions to be unreliable due to the significant difference in mean scores on the same passage. Amal Mahmoud (1977) in his masters thesis (cited in Rand, 1978, p.63) also found mean scores differed substantially, this time with deletions under 100. He recommended the use of 100 deletions as a minimum for cloze testing.

Sciarone and Schoorl (1989) in their study into the minimum number of deletions to ensure parallelism, also recommended 100 deletions for the exact word method of grading, but 75 deletions for the synonymic word method of grading. They conducted a study on groups of Indonesian learners of Dutch, aged between 17 and 21 years of age, who were seeking admission to the Delft University of Technology in The Netherlands. The experiment was aimed at determining the minimum number of deletions required to ensure parallelism for cloze tests differing at the point at which the deletion starts. They proposed a need for more deletions because of the possibility of the deletions falling in with the rhythm of the language and removing key elements from the passage. Longer passages will increase the possibility of the rhythm of the language breaking with the regular deletions (Taylor, 1953, p.419). They also argued that the accepted standard of a maximum of 50 word deletions for validity purposes, was not based on any experimental evidence (p.417).

They concluded in their report that this generally accepted maximum of 50 word deletions was insufficient to attain parallelism. In their opinion, 50 word deletions were not sufficient to allow the chances of mechanically selecting easy or hard words to cancel out

(p.428). Their study indicated that tests using the exact word scoring method should contain a minimum of 100 word deletions, with the synonymic scoring method requiring a minimum of 75 word deletions.

In conclusion, there have been a limited number of research studies into the number of deletions required to test comprehension through the use of cloze procedures. The research that has been completed recommends between 25 and 100 deletions, a range large enough to cause concern about the validity and reliability of cloze as a measure of comprehension. The accepted standard of 50 deletions proposed by Taylor (1953) would appear to be seriously questioned by the results of Sciarone et al., (1989) yet this research has some limitations. The results of the study are limited to adult Indonesian speakers of Dutch and the question remains as to whether these results could be obtained with English speaking Australian children. There is also a serious flaw in the research design. A 200 word deletion cloze test was administered and graded after various subsets of 100, 75 and 50 deletions. A test constructed in this manner would allow the students to go back through the text and alter previous answers after reading on and obtaining further meaning. As the students were not administered the 100, 75 and 50 word deletion versions as separate tests the results are invalid. The only score that is valid is the result obtained from the 200 word deletion test.

The review of the literature shows that it is necessary for further research to be conducted into the number of deletions required to ensure valid and reliable cloze tests. The research conducted thus far is contradictory and inconclusive, only adding to the difficulty facing teachers when attempting to construct cloze comprehension tests in the classroom.

Summary

In this chapter, the literature pertaining to reading, comprehension and the use of cloze procedures was examined. In particular it examined the background to cloze procedure, different construction methods, the use of random deletions, cloze sensitivity to intersentential constraint, different methods of grading cloze tests and the number of deletions required in cloze tests. It was shown that there is conflicting evidence in the literature in relation to the validity and reliability of different cloze construction methods.

This conflicting evidence raised three major questions that this study addresses to assist teachers in their construction of valid and reliable cloze tests for use in the classroom.

1) How many deletions should be included to ensure valid and reliable results on different texts?

2) Are the validity and reliability of Cloze Tests affected by different deletion starting points?

3) Should Cloze Tests be graded with the exact word scoring method or the synonymic word scoring method?

This review of the literature and the questions raised form the basis of the theoretical rationale adopted for this study in which the practical aspect of cloze construction in the classroom is the primary concern.

Chapter 3

Theoretical Rationale

Cloze comprehension testing is based on the law of closure, a major concept in Gestalt psychology. This theory was developed by three German psychologists, Max Wertheimer, Wolfgang Kohler, and Kurt Koffka in the first half of this century. An English approximation of the term "gestalt" would be a combination of the terms "form", "figure", "configuration" and "overall pattern" (Soudek & Soudek, 1983). The law of closure "reflects the natural tendency of human beings to perceive unfinished or incomplete figures as completed entities, to fill the gaps in broken patterns" (Soudek et al., 1983, p.335).

This psychological perspective of cloze procedure has been widely accepted though not without some challenge. Ohnmacht, Weaver and Kohler (1970) (cited in Rye, 1982, p.3) concluded, there was no strong relationship between the gestalt principle of closure and the completion items of a cloze passage, their argument being, cloze is essentially a cognitive task and not just a process of completing patterns.

According to Rye (1982), the cloze procedure approximates the strategies used in the reading process because " when completing a cloze deletion the reader samples the context information, constructs a response and then checks this response with the available context information" (p.7). When reading an undeleted text, the use of these strategies takes place at a subconcious level, based on previous experiences of language and understandings of meaning. Successful completion of cloze deletions requires a concious search of the wider context to obtain meaning to predict the deleted word without the assistance of graphic clues (Rye, 1982).

Chapter 4

Methodology

<u>Design of the Study</u>

The structure of the study is a correlational design in which the standardised <u>Gap</u> <u>Reading Comprehension Test</u> was administered to Year Five primary school children followed by two 100 word deletion Cloze Comprehension Tests, each divided into four subtests. Each subtest used the fixed ratio deletion method, with every fifth word in the passages deleted, differing only in the starting point for deletions between the fifth and sixth words of the second sentence.

Cloze Test A was taken from the story `Kiya the Gull' by Fen H. Lassell, from level 12 of <u>The Holt Basis Reading System Special Happenings</u> (Evertts & Weiss, 1977). The four subtests taken from this story were:

Form A1   - 50 word deletions starting at the fifth word of the second sentence.

Form A2   - 50 word deletions starting at the sixth word of the second sentence.

Form A3   -100 word deletions starting at the fifth word of the second sentence.

Form A4   - 100 word deletions starting at the sixth word of the second sentence.

Cloze Test B was taken from the story `Hattie, the Backstage Bat' by Don Freeman, from level 12 of The Holt Basic Reading System Special Happenings (Evertts & Weiss, 1977). The four subtests taken from this story were:

Form B1 - 50 word deletions starting at the fifth word of the second sentence.

Form B2 - 50 word deletions starting at the sixth word of the second sentence.

Form B3 - 100 word deletions starting at the fifth word of the second sentence.

Form B4 - 100 word deletions starting at the sixth word of the second sentence.

The tests were designed in this manner for a number of reasons. The construction of separate 50 and 100 word deletion tests was necessary to ensure the results were not influenced by the students changing their responses in the first half of the test (the 50 word deletion section), after having read on and obtained further meaning from the second half of the test.

Two different forms of each 50 and 100 word deletion subtest were constructed, differing only in the starting point for the deletions at either the fifth or sixth word of the second sentence, to ascertain if the validity and reliability of the cloze tests would be influenced by varying the starting point for the deletions.

The variables of interest which were correlated in order to answer each research question are as follows:

Research question no.1.

Is the 50 word deletion method of Cloze Comprehension Testing a valid measure of reading comprehension when:

a) The exact word scoring method is used?

b) The synonymic word scoring method is used?

The Gap Reading Comprehension Test score was correlated with the scores obtained from all 50 word deletion subtests graded with the exact word scoring method (exact word = EX) - Cloze Tests A1EX, A2EX, B1EX, and B2EX.

The Gap Reading Comprehension Test score was correlated with the scores obtained from all 50 word deletion subtests graded with the synonymic word scoring method (synonymic word = SYN) - Cloze Tests A1SYN, A2SYN, B1SYN and B2SYN.

Research question no.2.

Is the 100 word deletion method of Cloze Comprehension Testing a valid measure of reading comprehension when:

a) The exact word scoring method is used?

b) The synonymic word scoring method is used?

The Gap Reading Comprehension Test score was correlated with the scores obtained from all 100 word deletion subtests graded with the exact word scoring method - Cloze Tests A3EX, A4EX, B3EX, and B4EX.

31

The Gap Reading Comprehension Test score was correlated with the scores obtained from all 100 word deletion subtests graded with the synonymic word scoring method - Cloze Tests A3SYN, A4SYN, B3SYN and B4SYN.

Research question no.3.

Is the 50 word deletion method of Cloze Comprehension Testing a reliable measure of reading comprehension when:

a) The exact word scoring method is used?

b) The synonymic word scoring method is used?

The scores obtained from the odd numbered items were correlated with the scores obtained from the even numbered items on all 50 word deletion Cloze Tests graded with the exact word scoring method - Cloze Tests A1EX, A2EX, B1EX, and B2EX.

The scores obtained from the odd numbered items were correlated with the scores obtained from the even numbered items on all 50 word deletion subtests graded with the synonymic word scoring method - Cloze Tests A1SYN, A2SYN, B1SYN and B2SYN.

Research question no.4.

Is the 100 word deletion method of Cloze Comprehension Testing a reliable measure of reading comprehension when:

a) The exact word scoring method is used?

b) The synonymic word scoring method is used?

32

The scores obtained from the odd numbered items were correlated with the scores obtained from the even numbered items on all 100 word deletion Cloze Tests graded with the exact word scoring method - Cloze Tests A3EX, A4EX, B3EX, and B4EX.

The scores obtained from the odd numbered items were correlated with the scores obtained from the even numbered items on all 100 word deletion subtests graded with the synonymic word scoring method - Cloze Tests A3SYN, A4SYN, B3SYN and B4SYN.

Research Question No.5.

To what extent do measures using the synonymic word method of cloze grading correlate with the exact word method of cloze grading?

The results obtained from all subtests starting deletions at the fifth word of the second sentence and graded with the exact word scoring method - Cloze Tests A1EX(50), B1EX(50), A3EX(100) and B3EX(100) - were correlated with the results obtained from all subtests starting deletions at the fifth word of the second sentence and graded with the synonymic word scoring method - Cloze Tests A1SYN(50), B1SYN(50), A3SYN(100) and B3SYN(100).

The results obtained from all subtests starting deletions at the sixth word of the second sentence and graded with the exact word scoring method - Cloze Tests A2EX(50), B2EX(50), A4EX(100) and B4EX(100) - were correlated with the results obtained from all subtests starting deletions at the sixth word of the second sentence and graded with the synonymic word scoring method - Cloze Tests A2SYN(50), B2SYN(50), A4SYN(100) and B4SYN(100).

33

Description of Instruments

The following is a description of all instruments used in the study.

Gap Reading Comprehension Test

The Gap Reading Comprehension Test is a standardised reading comprehension test first published in 1965 but revised in 1976. The revised tests, which were administered to 250 children, used the split-half method to calculate reliability. Reliability coefficients, on samples of children from three different age groups, ranged from 0.90 to 0.94 (McLeod, 1977). There are two forms, red form R3 and blue form B3, which provide equivalent measures of reading comprehension. Form R3 contains 426 words, divided into eight unrelated paragraphs, with 43 word deletions. The words have been deleted by means of variable ratio deletion where specific word types have been deleted, in this case mainly function words, with the number of words between each deletion varying between seven and eleven words. Three of the paragraphs have the first sentence left intact while no paragraph has the last sentence left intact. Form B3 contains 412 words, divided into seven unrelated paragraphs, with 42 word deletions. Again, mainly function words have been deleted by means of variable ratio deletion, with the number of words between each deletion varying between six and eleven words. Three of the paragraphs have the first sentence left intact while no paragraph has the last sentence left intact.

Gap is a timed test in which the students are given 15 minutes to complete the test. Only the exact word replacement is scored as correct; synonyms are not scored as correct. As Gap is not a test of spelling ability incorrectly spelled versions of the correct response are scored as correct.

34

Cloze Tests

Cloze Tests prefixed A, `Kiya the Gull' by Fen H. Lasell and Cloze Test prefixed B, `Hattie the Backstage Bat' by Don Freeman, are both stories contained in the publication Special Happenings (Evertts & Weiss, 1977). This is a level 12 book of The Holt Basic Reading System which equates to the upper Year 5 level of reading ability in Australia.

Two readability formulae were applied to both stories, the Fry Readability Formula and the Readability Index (RIX). The RIX, determined by dividing the number of words with seven characters or more by the number of sentences, produced scores of .93 and 1.21 for Cloze Test A and B respectively. These scores both fell in the Grade 4 level on the RIX equivalent grade level table (Australia).

The Fry Readability Formula, determined by plotting the average syllable count and the average sentence count, in three 100 word selections, on a graph provided by Fry, produced grade levels of 3 and 5 for Cloze Tests A and B repectively (U.S. norms). The results of the readability formulae, though only estimations of text suitability, deemed the texts suitable for the middle primary level.

Cloze Test A was then pilot tested on a Year 5 class to use the cloze procedure as a means of judging readability. The scale used to determine readability was the criteria developed by Bormuth (1968, cited in Lipson et al., 1991, p.408). A percentage cloze score for the test was obtained by dividing the total number of correct responses for the children, by the total number of possible items. This produced a percentage score to be interpreted in the following manner:

Above 57%   Independent Reading Level

35

44-57%   Instructional Reading Level

Below 44%   Frustrational Reading Level

The results of the pilot test produced a 47% cloze score, which is at the instructional reading level, a level where the children would be able to cope with most of the reading but would need assistance to gain a deeper understanding of the text (Rye, 1982, p.22). This text level was deemed suitable as it would extend the abilities of the children at this level providing a greater insight into just what the Cloze Test was measuring.

The varying results of the readability formulae serve to confirm the caution expressed earlier in assigning texts to children by use of such formulae.

Reliability

Reliability of the Cloze Tests was determined by means of the split-half method. This method measures the internal consistency of the test. The items on the tests were divided into odd and even items to give each subject two scores, a score for the odd items and a score for the even items. The two scores were then correlated. If the coefficient is high then the test has good split-half reliability.

Validity

Concurrent validity of the Cloze Tests was established by means of correlating the scores from the Cloze Tests with the scores from the Gap Test. This is the degree to which scores on a test are related to the scores on an already established test (Gay, 1985, p.161). The degree of relationship between the two variables is expressed as a correlation coefficient.

Two university lecturers with specialisations in the area of reading were interviewed to determine content validity of Cloze Comprehension Tests. This focused on four particular areas:

1) Cloze as a measure of comprehension.

2) Fixed ratio and rational deletions.

3) Number of deletions.

4) Grading with the synonymic word and exact word methods.

In their opinion cloze as a measure of comprehension was largely dependent on the type of words that had been deleted. They preferred the use of rational deletions as it allowed the deletion of particular words in relation to the surrounding context and, deleting words in this manner, would be a good measure of text implicit strategies. Fixed-ratio deletions, which deleted a high ratio of function words, were deemed to be a measure of syntax rather than comprehension, allowing for completion of the test without full understanding. Opinions on the number of deletions varied with one lecturer advising 100 word deletions was a good idea after having first introduced the cloze procedure with 50 word deletions. Another was of the opinion that it was the ratio of deletions that was important rather than the number of deletions. Both were in agreement that the synonymic word scoring method was superior as there were varying degrees of correctness in reading comprehension. The use of synonymic scoring allowed the teacher to assess the child's developmental level and, over a period of time, build up a good picture of the child's comprehension ability by analysing the type of errors committed.

Subjects

Two schools from the Ministry of Education's classification six primary schools had originally been selected through stratified sampling for this study. Permission to

37

conduct the study in the chosen schools was denied by the principals on the grounds that they received numerous requests to conduct research in their schools and the teachers felt this was disruptive to the children's education. Two suburban primary schools were then selected on the basis of their availability. The number of subjects in the study was dictated by the total number of Year 5 students in both schools. Written permission was then obtained from the parents of each student.

This resulted in a total of 49 subjects, 27 from School A and 22 from School B, ranging in age from 9 years 9 months to 10 years 8 months. Three of the children were bilingual with English being the language predominantly spoken at home.

Data Collection Procedures

The data was collected by means of three different tests. The Gap Reading Comprehension Test and two 100 word deletion Cloze Comprehension Tests: Cloze Test A - `Kiya the Gull' and Cloze Test B - `Hattie, the Backstage Bat'. The Gap Reading Comprehension Test was administered to the children in Week 1 and the Cloze Tests to the children in Week 2. There were absentees from each of the sessions, which resulted in some of the children receiving only one of the tests which invalidated their data and were not included.

Two parallel forms of the Gap Reading Comprehension Test, blue form B3 and red form R3, were administered alternatively to the children in each class. This resulted in 23 children receiving Gap Form R3 and 23 children receiving Gap Form B3.

For the administration of the two Cloze Tests each class was divided into two halves, by means of every second student in their seating arrangements, resulting in four overall groups. The groups received the following Cloze subtests:

Group 1 - School 1 - subtests A1 then B4 - 13 and 14 students respectively.

Group 2 - School 1 - subtests A3 then B2 - 14 and 12 students respectively.

Group 3 - School 2 - subtests B3 then A2 - 10 students for each test.

Group 4 - School 2 - subtests B1 then A4 - 12 and 11 students respectively.

The format for the presentation of the Cloze Tests is presented below.

Test 1

Deletions Beginning at the 5th Word

Group No.	1	2	3	4
Test No.	A1	A3	B3	B1
Deletions	50	100	100	50

Test 2

Deletions Beginning at the 6th Word

Group No.	1	2	3	4
Test No.	B4	B2	A2	A4
Deletions	100	50	50	100

The difference in the number of students for each test was due to some of the students not completing the second test. While there was no set time limit for the completion of the tests, it became evident that some of the students were struggling to complete the second test and these tests were withdrawn for data analysis purposes.

40

The order of presentation was counterbalanced to eliminate any affects that may have occurred due to receiving the 50 word deletion test before the 100 word deletion test and vice - versa.

The tests were graded using both the exact word and synonymic scoring methods. A synonymic scoring key was constructed by presenting cloze subtests A3, A4, B3, and B4 (100 word deletion tests beginning at the fifth and sixth word) to qualified teachers who were asked to insert as many suitable responses they thought were possible for each blank space. This resulted in 12 responses for subtest A3, 10 responses for subtest A4 and 9 responses each for subtests B3 and B4. Only synonyms contained in the scoring key were scored as correct.

As with the Gap Test, the Cloze Test was not a measure of spelling ability and incorrect spelling of correct responses, exact word or synonymic word, were scored as correct.

Data Analysis Procedures

The data were analysed by means of correlation matrices, using the Pearson Product-Moment Correlation Coefficient using Statistical Analysis Software (SAS).

Chapter 5

Results

The results are analysed by first of all presenting the descriptive statistics, followed by validity coefficients in answer to research questions 1 and 2, then reliability coefficients in answer to research questions number 3 and 4. Finally, the exact word and synonymic word coefficients are presented in answer to research question number 5.

Analysis of Results

Table 5.1 presents the descriptive statistics for the 50 and 100 word deletion Cloze Tests which were graded using the exact word scoring method (EX). It is particularly noteworthy that the standard deviation of Cloze Test A3 (100) exact word (10.6) was substantially higher than any other test. This could be explained by the extreme low score of one particular student (23). Also of note is the low mean score of Cloze Test A2 (50) exact word (18.3) compared with the mean score of the other test that group of students completed B3 (100) exact word (49.2). Cloze Test B1 (100) exact word has a low distribution of scores with a standard deviation of 2.7.

Descriptive Statistics

Table 5.1

Descriptive Statistics for Cloze Tests Graded with the Exact Word

Scoring Method.

Variable	N	Mean	% Correct	SD	Range Actual	Possible
A1EX * (50)	13	26	52	4.8	19-35	0- 50
A2EX ** (50)	10	18.3	36.6	6.6	9-28	0- 50
B1EX * (50)	12	21.5	43	2.7	18-28	0- 50
B2EX ** (50)	12	24.8	49.6	6.4	9-33	0- 50
A3EX * (100)	14	46.2	46.2	10.6	23-60	0-100
A4EX ** (100)	11	48.1	48.1	4.6	38-54	0-100
B3EX * (100)	10	49.2	49.2	7.0	40-64	0-100
B4EX ** (100)	11	55.4	55.4	6.6	47-65	0-100

* - Every 5th word deletion beginning at the 5th word.

** - Every 5th word deletion beginning at the 6th word.

Table 5.2 presents the descriptive statistics for the 50 and 100 word deletion Cloze Tests which were graded using the synonymic word scoring method (SYN). The synonymic scores provide, as expected, a higher mean score than the exact word scoring method and again one student's low score affected the standard deviation on Cloze Test A3 (100) synonymic word. Cloze Test B1 (50) synonymic word score again presents a low distribution of scores (3.9).

Table 5.2

Descriptive Statistics for Cloze Tests Graded with the Synonymic Word Scoring Method.

Variable	N	Mean	% Correct	SD	Range Actual	Possible
A1SYN * (50)	13	38	76	6.5	27-47	0- 50
A2SYN ** (50)	10	32.7	65.4	7.5	21-42	0- 50
B1SYN * (50)	12	35	70	3.9	25-40	0- 50
B2SYN ** (50)	12	36.2	72.4	7.5	18-45	0- 50
A3SYN * (100)	14	75.1	75.1	13.5	52-89	0-100
A4SYN ** (100)	11	77.4	77.4	8.7	56-86	0-100
B3SYN * (100)	10	71.1	71.1	7.0	60-84	0-100
B4SYN ** (100)	11	81	81	8.3	67-91	0-100

* - Every 5th word deletion beginning at the 5th word.

** - Every 5th word deletion beginning at the 6th word.

Table 5.3 presents the descriptive statistics for the <u>Gap Reading Comprehension Test</u>.

<u>Table 5.3</u>

<u>Descriptive Statistics for the</u>

<u>Gap Reading Comprehension Test</u>

Variable	N	Mean	% Correct	SD	Range Actual	Possible
Gap	47	31.9	74.2	5.1	16-40	0-43

<u>Validity Coefficients</u>

<u>Research Question Number One</u>

Is the 50 word deletion method of Cloze Comprehension Testing a valid measure of reading comprehension when:

a) The exact word scoring method is used?

b) The synonymic word scoring method is used?

Concurrent validity was established by determining the degree of relationship between the scores on the Cloze Tests and the scores on the established standardised <u>Gap Reading Comprehension Test</u>.

The correlation Coefficients which were calculated between the Gap Test and the Cloze Tests containing 50 word deletions beginning at the fifth and sixth words and graded with the exact word scoring method are presented in Table 5.4.

Table 5.4 shows Cloze Tests A1EX and B2EX were significantly correlated with the Gap Test at the .05 level, Cloze Test A2EX was significantly correlated with the Gap Test at the .01 level, while Cloze Test B1EX was not significantly correlated.

Table 5.4

Correlation Coefficients Between the Gap Reading Comprehension

Test and Cloze Tests A1EX, A2EX, B1EX and B2EX.

	A1EX	A2EX	B1EX	B2EX
	(50)	(50)	(50)	(50)
	(5th)	(6th)	(5th)	(6th)
GAP	0.59*	0.77**	0.32	0.70*
N	13	10	12	12

* p<.05   ** p<.01

The correlation coefficients which were calculated between the Gap Test and the parallel forms of the Cloze Tests are presented in Table 5.5. The parallel forms correlation coefficients were determined in the following manner.

The scores obtained from Cloze Tests A1EX and B1EX , 50 word deletion tests starting deletions at the fifth word, were combined and named Cloze Test AB1EX. These results were then correlated with the Gap Test.

The scores obtained from Cloze Tests A2EX and B2EX, 50 word deletion tests starting deletions at the sixth word, were combined and named Cloze Test AB2EX. These results were then correlated with the Gap Test.

Table 5.5 shows Cloze Test AB2EX to be significantly correlated with the Gap Test at the .05 level while Cloze Test AB1EX was not significantly correlated.

Table 5.5

Correlation Coefficients Between the Gap Reading Comprehension
Test and Parallel Cloze Test Forms AB1EX and AB2EX.

	ABEX1 (0) (5th)	ABEX2 (50) (6th)
Gap	0.37	0.51*
N	23	22
* p<.05		

The correlation Coefficients which were calculated between the Gap Test and the Cloze Tests containing 50 word deletions beginning at the fifth and sixth words and graded with the synonymic word scoring method are presented in Table 5.6.

Table 5.6 shows that Cloze Tests A2SYN and B2SYN were significantly correlated with the Gap Test at the .01 level, Cloze Test A1SYN was significantly correlated with the Gap Test at the .001 level, while Cloze Test B1SYN was not significantly correlated.

Table 5.6

Correlation Coefficients Between the Gap Reading Comprehension
Test and Cloze Tests A1SYN, A2SYN, B1SYN and B2SYN.

	A1SYN	A2SYN	B1SYN	B2SYN
	(50)	(50)	(50)	(50)
	(5th)	(6th)	(5th)	(6th)
GAP	0.81***	0.80**	0.47	0.81**
N	13	10	12	12

** p<.01 *** p<.001

Research Question Number Two

Is the 100 word deletion method of Cloze Comprehension Testing a valid measure of reading comprehension when:

a) The exact word scoring method is used?

b) The synonymic word scoring method is used?

Concurrent validity was established by determining the degree of relationship between the scores on the Cloze Tests and the scores on the established standardised Gap Reading Comprehension Test.

The correlation coefficients which were calculated between the Gap Test and the Cloze Tests containing 100 word deletions beginning at the fifth and sixth words and graded with the exact word scoring method are presented in Table 5.7.

Table 5.7 shows Cloze Tests A3EX and B3EX were significantly correlated with the Gap Test at the .001 level, Cloze Test A4 was significantly correlated with the Gap Test at the .05 level, while Cloze Test 4EX was not significantly correlated.

Table 5.7

Correlation Coefficients Between the Gap Reading Comprehension Test and Cloze Tests A3EX, A4EX, B3EX and B4EX.

	A3EX (100) (5th)	A4EX (100) (6th)	B3EX (100) (5th)	B4EX (100) (6th)
GAP	0.88***	0.65*	0.88***	0.60
N	14	11	10	11

* p<.05 *** p<.001

The correlation coefficients which were calculated between the Gap Test and the parallel forms of the Cloze Tests are presented in Table 5.8. The parallel forms correlation coefficients were determined in the following manner.

The scores obtained from Cloze Tests A3EX and B3EX , 100 word deletion tests starting deletions at the fifth word, were combined and named Cloze Test AB3EX. These results were then correlated with the Gap Test.

The scores obtained from Cloze Tests A4EX and B4EX, 100 word deletion tests starting deletions at the sixth word, were combined and named Cloze Test AB4EX. These results were then correlated with the Gap Test.

Table 5.8 shows Cloze Test AB4EX to be significantly correlated with the Gap Test at the .05 level and Cloze Test AB3EX to be significantly correlated with the Gap Test at the 0.001 level.

Table 5.8

Correlation Coefficients Between the Gap Reading Comprehension
Test and Parallel Cloze Test Forms AB3EX and AB4EX.

	AB3EX (100) (5th)	AB4EX (100) (6th)
Gap	0.89***	0.43*
N	24	22
* p<.05 *** p<.001		

The correlation coefficients which were calculated between the Gap Test and the Cloze Tests containing 100 word deletions beginning at the fifth and sixth words and graded with the synonymic word scoring method are presented in Table 5.9.

Table 5.9 shows all Cloze Tests containing 100 word deletions and graded with the synonymic word scoring method to be significantly correlated with the Gap Test, Cloze Tests A4SYN and B4SYN at the .05 level and Cloze Tests A3SYN and B3SYN at the .01 level.

Table 5.9

Correlation Coefficients Between the Gap Reading Comprehension Test and Cloze Tests A3SYN, A4SYN, B3SYN and B4SYN.

	A3SYN (100) (5th)	A4SYN (100) (6th)	B3SYN (100) (5th)	B4SYN (100) (6th)
GAP	0.78**	0.65*	0.85**	0.71*
N	14	11	10	11

* p<.05 ** p<.01

Research Question Number Three

Is the 50 word deletion method of Cloze Comprehension Testing a reliable measure of reading comprehension when:

a) The exact word scoring method is used?

b) The synonymic word scoring method is used?

The reliablity of the Cloze Tests was examined using the split-half method to determine internal consistency by dividing the tests into the two comparable halves by means of odd and even numbered items.

The internal consistency correlation coefficients which were calculated between odd and even numbered items on Cloze Tests containing 50 word deletions beginning at the fifth and sixth words and graded with the exact word scoring method are presented in Table 5.10.

Table 5.10 shows that no test was reliable and that there was a negative correlation for Cloze Test B1EX with practically no relationship between the halves of the test.

54

Table 5.10

Internal Consistency Coefficients Between Odd and Even numbered
Items on Cloze Tests A1EX, A2EX, B1EX and B2EX.

	A1EX	A2EX	B1EX	B2EX
	(50)	(50)	(50)	(50)
	(5th)	(6th)	(5th)	(6th)
	0.43	0.55	-0.013	0.22
N	13	10	12	12

The correlation coefficients which were calculated between odd and even numbered items on the parallel forms of the Cloze Tests are presented in Table 5.11. The parallel forms correlation coefficients were determined in the following manner.

The scores obtained from Cloze Tests A1EX and B1EX , 50 word deletion tests starting deletions at the fifth word, were combined and named Cloze Test AB1EX. The odd and even numbered items on this test were then correlated.

The scores obtained from Cloze Tests A2EX and B2EX, 50 word deletion tests starting deletions at the sixth word, were combined and named Cloze Test AB2EX. The odd and even numbered items on this test were then correlated.

55

Table 5.11 shows both parallel forms to be reliable being significantly correlated at the .05 level.

Table 5.11

Internal Consistency Coefficients Between Odd and Even numbered
Items on Parallel Form Cloze Tests AB1EX and AB2EX.

	AB1EX (50) (5th)	AB2EX (50) (6th)
	0.42*	0.51*
N	25	22

* p<.05

The internal consistency correlation coefficients which were calculated between odd and even numbered items on Cloze Tests containing 50 word deletions beginning at the fifth and sixth words and graded with the synonymic word scoring method are presented in Table 5.12.

Table 5.12 shows that three of the four tests were reliable. Cloze Test B1SYN was significantly correlated at the .05 level, Cloze Test A2SYN was significantly correlated at the .01 level and Cloze Test A1SYN was significantly correlated at the .001 level.

Table 5.12

Internal Consistency Coefficients Between Odd and Even numbered Items on Cloze Tests A1SYN, A2SYN, B1SYN and B2SYN.

	A1SYN (50) (5th)	A2SYN (50) (6th)	B1SYN (50) (5th)	B2SYN (50) (6th)
	0.83***	0.84**	0.59*	0.30
N	13	10	12	12

* p<.05 ** p<.01 *** p<.001

Research Question Number Four

Is the 100 word deletion method of Cloze Comprehension Testing a reliable measure of reading comprehension when:

a) The exact word scoring method is used?

b) The synonymic word scoring method is used?

The internal consistency correlation coefficients which were calculated between odd and even numbered items on Cloze Tests containing 100 word deletions beginning at the fifth and sixth words and graded with the exact word scoring method are presented in Table 5.13.

57

Table 5.13 shows that three of the four tests were reliable. Cloze Tests B3EX and B4EX were significantly correlated at the .05 level and Cloze Test A3EX was significantly correlated at the .001 level.

Table 5.13

Internal Consistency Coefficients Between Odd and Even numbered Items on Cloze Tests A3EX, A4EX, B3EX and B4EX.

	A3EX	A4EX	B3EX	B4EX
	(100)	(100)	(100)	(100)
	(5th)	(6th)	(5th)	(6th)
	0.84***	0.44	0.77*	0.59*
N	14	11	10	11

* p<.05 *** p<.001

The correlation coefficients which were calculated between odd and even numbered items on the parallel forms of the Cloze Tests are presented in Table 5.14. The parallel forms correlation coefficients were determined in the following manner.

The scores obtained from Cloze Tests A3EX and B3EX , 100 word deletion tests starting deletions at the fifth word, were combined and named Cloze Test AB3EX. The odd and even numbered items on this test were then correlated.

The scores obtained from Cloze Tests A4EX and B4EX, 100 word deletion tests starting deletions at the sixth word, were combined and named Cloze Test AB4EX. The odd and even numbered items on this test were then correlated.

Table 5.14 shows only parallel form AB3EX to be reliable being significantly correlated at the .001 level.

<u>Table 5.14</u>

<u>Internal Consistency Coefficients Between Odd and Even numbered</u>

<u>Items on Parallel Form Cloze Tests AB3EX and AB4EX.</u>

	AB3EX (100) (5th)	AB4EX (100) (6th)
	0.67***	0.23
N	24	22

*** p<.001

The internal consistency correlation coefficients which were calculated between odd and even numbered items on Cloze Tests containing 100 word deletions beginning at the fifth and sixth words and graded with the synonymic word scoring method are presented in Table 5.15.

Table 5.15 shows that all four tests were reliable. Cloze Test A4SYN was significantly correlated at the .05 level, Cloze Tests B3SYN and B4SYN were significantly correlated at the .01 level and Cloze Test A3SYN was significantly correlated at the .001 level.

Table 5.15

Internal Consistency Coefficients Between Odd and Even numbered
Items on Cloze Tests A3SYN, A4SYN, B3SYN and B4SYN.

	A3SYN	A4SYN	B3SYN	B4SYN
	(100)	(100)	(100)	(100)
	(5th)	(6th)	(5th)	(6th)
	0.99***	0.71*	0.80**	0.73**
N	14	11	10	11

* p<.05 ** p<.01 *** p<.001

60

Research Question No. 5

To what extent does the synonymic word method of cloze grading correlate with the exact word method of cloze grading?

The correlation coefficients which were calculated between all Cloze Tests A graded with the exact word scoring method and all Cloze Tests A graded with the synonymic word scoring method are presented in Table 5.16.

Table 5.16 shows all exact word and synonymic word tests were significantly correlated at the .001 level. Correlation coefficients were high, ranging from 0.82 to 0.92.

Table 5.16

Cloze Tests A Graded with the Exact Word Scoring Method and Cloze
Tests A Graded with the Synonymic Word Scoring Method Correlation
Coefficients.

	A1SYN (50/5th)	A2SYN (50/6th)	A3SYN (100/5th)	A4SYN (100/6th)
A1EX (50/5th)	0.82***			
N	13			
A2EX (50/6th)		0.87***		
N		10		
A3EX (100/5th)			0.92***	
N			14	
A4EX (100/6th)				0.88***
N				11

*** p<.001

The correlation coefficients which were calculated between all Cloze Tests B graded with the exact word scoring method and all Cloze tests B graded with the synonymic word scoring method are presented in Table 5.17.

Table 5.17 shows all exact word and synonymic word tests were significantly correlated, Cloze Test B1 at the .01 level and Cloze Tests B2, B3 and B4 at the .001 level. Cloze Test B1 had a moderately low correlation coefficient of 0.71 with all other tests ranging from 0.89 to 0.93.

Table 5.17

Cloze Tests B Graded with the Exact Word Scoring Method and Cloze Tests B Graded with the Synonymic Word Scoring Method Correlation Coefficients.

	B1SYN (50/5th)	B2SYN (50/6th)	B3SYN (100/5th)	B4SYN (100/6th)
B1EX (50/5th)	0.71**			
N	12			
B2EX (50/6th)		0.92***		
N		12		
B3EX (100/5th)			0.93***	
N			10	
B4EX (100/5th)				0.89***
N				11

** p<.01   *** p<.001

The ranked order of the subjects from School No.1 when the Cloze tests are graded with both the exact word and synonymic word scoring methods is presented in Table 5.18. In order to achieve a discernible difference in ranked order the subjects have been ranked by their total scores from both the 50 word deletion and 100 word deletion Cloze Tests. Only those subjects who completed both tests have been included.

With the exception of subjects no.14 and no.17 (down 6 places on synonymic score) and subjects no.9 and no.16 (up 4 places on synonymic score) Table 5.18 shows there is no discernible difference in the subject's ranked order when the tests are graded with either the exact word or synonymic word scoring methods.

Ranked Order

Table 5.18

Ranking of Subjects From School No.1 According to Their Exact

Word and Synonymic Word Scores.

Rank	Subject	Exact Word Score	Subject	Synonymic Word Score
1	1	32	1	70
2	2	50	2	82
3	3	66	3	98
4	4	66	7	106
5	5	70	4	107
6	6	70	8	107
7	7	71	6	108
8	8	71	14	114
9	9	77	10	114
10	10	77	11	115
11	11	78	17	117
12	12	79	13	122
13	13	80	9	124
14	14	82	15	125
15	15	83	12	126
16	16	84	5	127
17	17	85	21	127
18	18	86	19	131
19	19	87	18	132
20	20	90	16	132
21	21	93	22	138
22	22	96	20	138

The ranked order of the subjects from School No.2 when the Cloze tests are graded with both the exact word and synonymic word scoring methods is presented in Table 5.19. In order to achieve a discernible difference in ranked order the subjects have been ranked by their total scores from both the 50 word deletion and 100 word deletion Cloze Tests. Only those subjects who completed both tests have been included.

With the exception of subjects no. 6 and no. 11 (up 6 and 7 places on synonymic score respectively) and subject no. 9 (down 4 places on synomymic score, Table 5.19 shows there is no discernible difference in the subject's ranked order when the tests are graded with either the exact word or synonymic word scoring methods.

Table 5.19

Ranking of Subjects From School No. 2 According to Their Exact Word and Synonymic Word Scores.

Rank	Subject	Exact Word Score	Subject	Synonymic Word Score
1	1	55	4	87
2	2	56	2	90
3	3	56	1	91
4	4	58	3	94
5	5	59	9	96
6	6	65	5	96
7	7	65	8	98
8	8	65	10	108
9	9	67	7	109
10	10	68	13	109
11	11	68	12	110
12	12	69	6	112
13	13	69	14	113
14	14	70	17	114
15	15	72	15	116
16	16	74	16	119
17	17	75	20	120
18	18	76	11	121
19	19	80	18	121
20	20	84	19	121
21	21	88	21	126

Chapter 6

Discussion and Conclusions

The purpose of the study is to provide guidelines to classroom teachers for the construction of valid and reliable Cloze Comprehension Tests for the practical application in the classroom. This chapter discusses the implications of the results for each research question and conclusions are drawn.

During the course of this chapter the answers to the research questions will be addressed in the following manner. First of all, in response to Research Questions 1-4, the validity and reliability of the 50 and 100 word deletion Cloze Tests when graded with the exact word scoring method will be discussed. This is followed by a discussion of the validity and reliability of the 50 and 100 word deletion Cloze Tests when graded with the synonymic word scoring method. Finally, in response to research question no. 5, the relationship between the exact word scoring method and the synonymic word scoring method will be discussed.

Research Question No. 1

Is the 50 word deletion method of Cloze Comprehension Testing a valid measure of reading comprehension when:

    a) the exact word scoring method is used?

    b) the synonymic word scoring method is used?

Research Question No. 2

Is the 100 word deletion method of Cloze Comprehension Testing a valid measure of reading comprehension when:

    a) the exact word scoring method is used?

    b) the synonymic word scoring method is used?

Research Question No. 3

Is the 50 word deletion method of Cloze Comprehension Testing a reliable measure of reading comprehension when:

    a) the exact word scoring method is used?

    b) the synonymic word scoring method is used?

Research Question No. 4

Is the 100 word deletion method of Cloze Comprehension Testing a reliable measure of reading comprehension when:

    a) the exact word scoring method is used?

    b) the synonymic word scoring method is used?

50 Word Deletion Cloze Tests

Exact Word Scoring Method

Concurrent validity was established for three of the four 50 word deletion Cloze Tests, A1EX (5th) (0.59, $p<.05$), A2EX (6th) (0.77, $p<.01$) and B2EX (6th) (0.70, $p<.05$) presented in Table 5.4, in that they were significantly correlated with the Gap Reading Comprehension Test. Cloze Test B1EX (5th) was not significantly correlated with the Gap Test. In order to determine a possible reason for Cloze Test B1EX not being significantly correlated with the Gap the mean Gap Test score of the group who completed Cloze Test B1EX was investigated. This showed that this particular group had the highest mean score on the Gap Test (34). An analysis of the types of words deleted in the Cloze Tests was then undertaken. This showed Cloze Test B1EX had 34 (68%) content words and 16 (32%) function words deleted from the passage. This is in contrast to the B2EX version of the test which had 28 (56%) content words and 22 (44%) function words deleted from the passage. As Alderson (1980) stated, the difference in the cloze test scores may be due to the difference in the particular words deleted from the passage. An analysis of the Gap Reading Comprehension Test revealed that 70% of the words deleted from the passages were function words. This could be a possible explanation for the insignificant relationship between this test (B1EX) and the Gap Test. As the Gap Test contains 43 deletions, it would therefore be expected that a 50 word deletion Cloze Test with a high degree of function words deleted would have a higher correlation with the Gap Test.

Reliability coefficients presented in Table 5.10 show the internal consistency of the 50 word deletion exact word tests to be very low. Correlation coefficients ranged from -0.013 (B1EX) to 0.55 (A2EX) with no split-half versions of the tests being significantly correlated.

71

The scores obtained on the 50 word deletion Cloze Tests beginning at the fifth word (A1EX and B1EX) and the sixth word (A2EX and B2EX) were then combined (AB1EX and AB2EX respectively) and correlated with the Gap Reading Comprehension Test to assess the concurrent validity of the parallel forms. This procedure was performed in order to discover whether the low reliability results could be attributed to the small number of subjects who took each parallel form of the test. This showed only parallel form AB2EX to be significantly correlated with the Gap Test (0.51, p<.05) see Table 5.5. Analysis of the types of words deleted showed parallel form AB2EX had 44% function words deleted while parallel form AB1EX had 40% function words deleted. Split-half reliability of the parallel forms, presented in Table 5.11, showed both parallel forms proved to be significantly correlated at the .05 level with coefficients of 0.42 (AB1EX) and 0.51 (AB2EX).

100 Word Deletion Cloze Tests

Exact Word Scoring Method

Concurrent validity was established for Cloze Tests A3EX (5th) (0.88, $p<.001$), A4EX (6th) (0.65, $p<.05$) and B3EX (5th) (0.88, $p<.001$). Cloze Test B4EX (6th) was not significantly correlated with the Gap Test. The two tests with the highest percentage of function words deleted (A3EX 49% and B3EX 46%) showed the highest correlation with the Gap Test. Split-half reliability coefficients, presented in Table 5.13, of three tests were significant; Cloze Tests A3EX (0.84, $p<.001$), B3EX (0.77, $p<.05$) and B4EX (0.59, $p.05$). The split half reliability coefficient for Cloze Test A4EX (40% function words deleted) was not significant. Parallel forms of the Cloze Tests presented in Table 5.8 show parallel forms AB3EX (A3EX and B3EX) (0.89, $p<.001$) and AB4EX (A4EX and B4EX) (0.43, $p<.05$) to be significantly correlated with the Gap Reading Comprehension Test.

Split-half reliability of the the parallel forms, presented in Table 5.14, show only parallel form AB3EX (0.67, $p<.001$) to be reliable. This form had 47% function words deleted, while parallel form AB4EX, where split-half reliability was not significant, had only 41% of the function words deleted.

Conclusion

The results would appear to support Mahmoud (1977, cited in Rand, 1978, p.63) and Sciarone et al., (1989) who recommended a minimum of 100 word deletions for cloze tests to be reliable and contradict Bachman (1985) and Rand (1978) who found reliability could be attained by 30 and 25 deletions respectively. While concurrent validity was established for three

73

out of the four 50 word deletion Cloze Tests (A1EX, A2EX and B2EX) reliability was not. Reliability was only established on three of the four 100 word deletion versions of the Cloze Tests (A3EX, A4EX and B3EX). This would appear to support Sciarone et al., (1989) who contended that there was a need for a large number of deletions to remove the possibility of the deletions falling in with the rhythm of the language and removing key elements from the passage. Taylor (1953, p.419) was also of this opinion stating that longer passages would increase the prospect of the rhythm of the language breaking with the regular deletions. This was particularly evident in Cloze Test B1EX where 68% of content words and 32% of function words were deleted in this 50 word deletion test. In the 100 word deletion version of this test (B3EX) the percentage of content and function words were deleted in a more balanced fashion, 54% content words and 46% function words. Cloze Test B3EX was significantly correlated with the Gap Reading Comprehension Test at the .001 level, perhaps because of the balance of deleted content and function words.

Alderson (1980) argued that random deletion ignored the syntactical-semantic relationship in the text and that the inconsistency in the results would be dependent on what proportion of syntactic and textual function words were deleted. As this study did not investigate this particular issue, it is impossible to know if the frequency of deleting function words was responsible for the inconsistency of results for some of the tests and the low split-half reliability of the 100 word deletion tests, but the results would indicate that these are possible explanations. It would appear that the closer the percentage of content and function words deleted is to a 50/50 ratio the greater the chances of achieving split-half reliability on the tests.

The results then indicate that 50 and 100 word deletion Cloze Tests, using random deletions and graded with the exact word scoring method, were in six out of eight cases, valid measures of comprehension. To increase the likelihood of achieving reliability, construct Cloze

Tests with 100 word deletions. This is necessary to ensure that the deletions do not fall in with the rhythm of the language and delete a high proportion of one particular type of word.

50 and 100 Word Deletion Cloze Tests

Synonymic Word Scoring Method

Concurrent validity was established for seven of the eight Cloze Tests graded with the synonymic word scoring method. These were 50 word deletion Cloze Tests A1SYN (0.81, $p<.001$), A2SYN (0.80, $p<.01$), and B2SYN (0.81, $p<.01$), presented in Table 5.6, and 100 word deletion Cloze Tests A3SYN (0.78, $p<.01$), A4SYN (0.65, $p<.05$), B3SYN (0.85, $p<.01$) and B4SYN (0.71, $p<.05$), presented in Table 5.9. This was achieved by correlating the scores on all the above tests with the scores on the Gap Reading Comprehension Test. The only Cloze Test graded with the synonymic scoring method that was not significantly correlated with the Gap Test was B1SYN (0.47). This test was the synonymic version of Cloze Test B1EX (0.32). As discussed earlier it seems that this test may be insignificantly correlated with the Gap Test because of the low ratio of function words deleted in the passage.

Split-half reliability was established for seven of the eight tests scored with the synonymic word method. These were 50 word deletion Cloze Tests A1SYN (0.83, $p<.001$), A2SYN (0.84, $p<.01$) and B1SYN (0.59, $p<.05$), presented in Table 5.12, and 100 word deletion Cloze Tests A3SYN (0.99, $p<.001$), A4SYN (0.71, $p<.05$), B3SYN (0.80, $p<.01$) and B4SYN (0.73, $p<.01$), presented in Table 5.15. The only Cloze Test where split-half reliability was not established was B2SYN.

It is interesting to note that three of the four 50 word deletion Cloze Tests, where split-half reliability was not established when graded with the exact word method, established split-half reliability when graded with the synonymic word method. Cloze Test B1 exemplified this with an internal consistency coefficient of -0.013 in the exact word version of the test and an internal consistency coefficient of 0.59 ($p<.05$) in the synonymic version of the test.

This pattern continued with the 100 word deletion versions with split-half reliability not established for Cloze Test A4 (0.44) when graded with the exact word method, but was established when graded with the synonymic word method (0.71, $p<.05$).

Conclusion

The arguments put forward for the use of exact word grading of cloze tests over the synonymic word scoring method focus on the objectivity of scoring when using the exact word scoring method (Alexander, 1968; Rankin, 1958; Smith and Zinc, 1977; Wiechelman, 1971, cited in Henk & Selders, 1984, p.282). While this is true, the results would suggest that there are other considerations when deciding on which method of grading to use.

Of the 8 Cloze Tests graded with the synonymic scoring method, 7 were shown to have a strong relationship with the Gap Reading Comprehension Test, with all correlation coefficients, except for B1SYN, being significant in a range of 0.65 to 0.85. Taylor (1956, p.48) suggested that there was no advantage to be gained by going to the trouble of judging and scoring synonyms and the subsequent research since has tended to support this view (Anderson, 1972; McKenna, 1976; Ruddell, 1964, cited in Henk, 1981, p.348). However, the results of this study do not support Taylor (1956). While it was true that in relation to the concurrent validity correlations with the Gap Reading Comprehension Test there was no

76

advantage to be gained, this was not the case with the split-half reliability of the tests. When the split-half reliability is taken into consideration, the results of this study show that there was indeed an advantage to be gained `by going to the trouble' of grading cloze tests with the synonymic scoring method. The split-half reliability of all tests was increased by the use of synonymic scoring. Four of the tests that did not show split-half reliability using the exact word method were reliable when graded with the synonymic word scoring method.

In summary, there is a high degree of relationship between cloze tests using the synonymic word scoring method and the Gap Reading Comprehension Test. When assessing the advantages of using the synonymic method purely in relation to its correlation with the Gap Test, the results would show no advantage in grading tests with the synonymic method. When taking the reliability of the tests into consideration, the results show a distinct advantage in using the synonymic method to achieve increased reliability.

Synonymic Versus Exact Word

Research Question No.5

To what extent do measures using the synonymic word method of cloze grading correlate with the exact word method of cloze grading?

Table 5.16 presents the correlation coefficients for all Cloze Tests prefixed A, which were taken from the story `Kiya the Gull', and graded with the exact word and synonymic word scoring methods. The results show a high degree of relationship between the synonymic word scoring method and the exact word scoring method. Correlation coefficients range from 0.82 to 0.92 with all tests significantly correlated ($p<.001$). Table 5.17 presents the correlation coefficients for all Cloze Tests prefixed B, which were taken from the story `Hattie the

Backstage Bat', and graded with the exact word and synonymic word scoring methods. Again the results show a high degree of relationship between the synonymic word and exact word scoring method. Correlation coefficients range from 0.71 (B1EX, $p<.01$) to 0.93 with Cloze Tests B2EX/SYN, B3EX/SYN and B4EX/SYN significantly correlated at the .001 level.

The ranked orders of the subjects from both schools according to their synonymic and exact word scores are presented in Tables 5.18 and 5.19. These tables show that 7 out of 43 students, subjects no.14 and no.17 (down 6 places on synonymic score) and subjects no.9 and no.16 (up 4 places on synonymic score) in Table 5.18, and subjects no.6 and no.11 (up 6 and 7 places respectively on synonymic score) and subject no.9 (down 4 places on synonymic score) in Table 5.19, had their ranked order changed as a result of grading the tests with the exact word and synonymic word scoring methods. It is assumed that the lower ranking of certain students on the tests graded with the synonymic word method, was due to a general increase in the other student's ability to complete the test when appropriate synomyms were scored as correct.

Conclusion

The results show that there is a close relationship between Cloze Tests graded with the exact word scoring method and the synonymic word scoring method. The close relationship would suggest there is little advantage in grading tests by the synonymic word method and support the evidence for the use of the exact word scoring method for ease and objectivity in marking the tests. (Alexander (1968); Rankin, 1968; Smith and Zinc, 1977; Taylor, 1956; Wiechelman, 1971, cited in Henk and Selders, 1984, p.282). Bachman (1967, cited in Rye, 1982, p.20) also stated the case for exact word scoring suggesting that the marking of synonyms would not significantly affect the results if the cloze pasages were long enough.

The consistency of the ranked order of the subjects in both conditions in this study would also support the research of Anderson (1972), Hargis (1972), McKenna (1976), and Ruddell (1964, cited in Henk, 1981, p.348) who found no appreciable differences in rank were obtained through synonymic scoring.

Summary and Recommendations

Three out of the four 50 word deletion Cloze Tests and three out of the four 100 word deletion Cloze Tests, graded with the exact word scoring method, were shown to be valid measures of reading comprehension in that they were significantly correlated with the Gap Reading Comprehension Test a widely used standardised reading test. All four of the 50 word deletion versions of the exact word tests were shown to be deficient in split-half reliability, which was achieved on three out of the four 100 word deletion exact word tests. This would suggest that while neither 50 nor 100 word deletion Cloze Tests can guarantee validity or reliability when graded with the exact word method, the 100 word deletion Cloze Tests were shown to be the more reliable measure of reading comprehension.

As with the exact word versions, the 50 word deletion synonymic versions of the Cloze Tests showed three out of the four tests to be valid measures of reading comprehension when correlated with the Gap Reading Comprehension Test. All four 100 word deletion Cloze Tests graded with the synonymic word scoring method were shown to be valid measures of reading comprehension. Three out of the four 50 word deletion Cloze Tests were shown to be have split-half reliability, while this was achieved in all four of the 100 word deletion Cloze Tests graded with the synonymic word scoring methods. This would suggest that while reliability can be achieved using 50 word deletion Cloze Tests, graded with the synonymic word scoring method, the 100 word deletion method, graded with the synonymic word scoring method, is the more reliable measure of reading comprehension.

The low percentage of function words deleted in some of the tests appeared to cause some inconsistencies in the results of both forms of the tests, reinforcing Alderson's (1980) argument, that the differences in cloze test scores may not be due to differences in deletion frequency but to differences in the particular words deleted. The use of 100 word deletions

80

appeared to correct this imbalance supporting Taylor's (1953) contention that the longer the passage the greater the prospect of the deletions breaking with the rhythm of the language, and also supports the the views of Sciarone et al., (1989) who recommended 100 word deletions for cloze tests to be reliable. As this study did not directly address the question of the types of words deleted from a passage, future research should address this issue to determine if the differences in cloze scores are due to the types of words deleted.

This study has limited reading comprehension to the children's ability to complete the cloze procedure and the results should be assessed in this context. The conclusions about the validity of the tests are limited to the correlation with the Gap Reading Comprehension Test. Further research should replicate this study with other standardised reading comprehension tests such as the Progressive Achievement Test (PAT) or the Torch Tests of Reading Comprehension to ascertain if these results would be reproduced.

The results also showed that the synonymic word method of cloze grading correlated highly with the exact word method of cloze grading. This would appear to confirm previous research which suggests there is no advantage to be gained from grading cloze tests with the synonymic word method (Taylor, 1956; Bachman, 1967, cited in Rye, 1982, p.20) and support the arguments for the use of the exact word method (Alexander, 1968; Rankin, 1968; Smith and Zinc, 1977; Wiechelman, 1971, cited in Henk and Selders, 1984 p.282). There was a small variation in the ranked order of the subjects between the synonymic and exact word scoring methods, supporting the research of Anderson (1972), Hargis (1972), McKenna (1976) and Ruddell (1964, cited in Henk, 1981, p.348).

The basis of the argument for the exact word scoring method focuses on the elimination of subjectivity in marking and the variability of an individual's cloze score (Henk et al., 1984, p.284). The results of this study indicate that the reliability of the cloze tests should

81

also be taken into consideration. The split-half reliability of all the Cloze Tests in this study increased with synonymic scoring. A synonymic scoring key was established for the Cloze Tests in this study by presenting the tests to qualified teachers for completion. While this method would be impractical for classroom use, in agreement with Anderson (1972), Bormuth (1965), Miller and Coleman (1967, cited in Henk et al., 1984, p.286) and Henk (1981), it is recommended that responses that preserve meaning should be counted.

Implications for Classroom Teachers.

The previous discussion of findings suggest that a reasonably reliable and valid test of comprehension can be constructed by using the cloze method. Even with validity and reliability established on a test, Rye (1982) suggested that teachers should analyse the results with three questions in mind.

1) What was the purpose behind my using the test?

2) Do the results show me particular areas of strengths and weaknesses?

3) How am I going to direct my teaching to help the child's needs?

An analysis of cloze test results should not just become a simple assignment of a particular score to a particular student. The scores obtained from the use of cloze procedures, as with any reading tests, should be treated with caution. With a cloze score of 60% or a standardised reading test age of 9 years 11 months, it is only possible to say that on a particular day, on a particular test, in a particular context, a child obtained a score of 60% or a reading age equivalent to 9 years 11 months. The results of any test can be influenced by a variety of extraneous variables, context, reader and text, and this should be taken into consideration in any analysis.

On the basis of the previous findings and discussion, the following guidelines are recommended for random deletion tests in which the deletions are made every fifth word throughout the passage.

1) In order to increase validity and prevent the deletions falling in with the rhythm of the language, and deleting key elements from the passage, 100 word deletions are recommended.

2) In order to ensure the reliability of cloze tests the synonymic word scoring method should be employed. It is advised that this should be achieved through increased training and guidelines for classroom teachers when grading synonyms, such as judging the synonym at the sentence level in regard to the context of the passage and not evaluating the answer at the word level only.

3) The use of the synonymic scoring method will render the criteria for judging cloze scores (frustration level, below 44%, instructional level, 44-57% and independent level, above 57%) developed by Bormuth (1968, cited in Lipson et al., 1991 p.408) invalid as they were based on the exact word scoring method.

4) Teachers should have a clear purpose for testing and should exhibit caution in their interpretation of a test score.

5) It is recommended that cloze tests be used in the classroom as a diagnostic tool rather than a general measure of reading comprehension ability. This diagnostic approach should focus on an analysis of the type of errors made in reference to their semantic and syntactic acceptability.

References

Alderson, J.C. (1979). The effect on the cloze test of changes in the deletion frequency. Journal of Research in Reading, 2 (2), 108-119.

Alderson, J.C. (1980). Native and non-native speaker performance on cloze tests. Language Learning, 30 (1), 59-76.

Bachman, L.F. (1982). The trait structure of cloze test scores. TESOL Quarterly, 16 (1), 61-70.

Bachman, L.F. (1985). Performance on cloze tests with fixed- ratio and rational deletions. TESOL Quarterly, 19 (3), 535-536.

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

Bormuth, J.R. (1969). Factor validity of cloze tests as measures of reading comprehension ability. Reading Research Quarterly, 4 (3), 358-365.

Burnes, D. (1985). Comprehending Text. In D. Burnes and G. Page, (eds.) Insights and Strategies for Teaching Reading. Sydney: Brace Jovanovich, pp.45-86.

De Santi, R.J. & Sullivan, V.G. (1984). Inter-rater reliability of the cloze reading inventory as a qualitative measure of reading comprehension. Reading Psychology, 5 (3-4), 203-208.

85

De Santi, R.J. & Sullivan, V.G. (1985). Reliability of single rater judgements of semantic and syntactic classifications of cloze test responses. Journal of Research and Development in Education, 18 (2), 49-53.

Duffelmeyer, F.A. (1983). The effect of grade level on cloze test scores. Journal of Reading, 26 (5), 436-441.

Evertts, E.L.& Weiss, B.J.(Eds.). (1977). Special Happenings (2nd Ed.). New York: Holt, Rinehart and Winston.

Ferguson, G.A. (1981). Statistical Analysis in Psychology and Education.(5th ed.) New York: McGraw-Hill.

Freedle, R.O.& Carroll, J.B.(Eds.). (1972). Language Comprehension and the Acquisition of Knowledge. Washington, D.C.: V.H. Winston & Sons.

Gay, L.R. (1981). Educational Research.(2nd ed.) Columbus, Ohio: Charles E. Merrill.

Gay, L.R. (1985). Educational Evaluation and Measurement. (2nd. ed.). Columbus, Ohio: Charles E. Merrill.

Gilliland, J. (1972). Readability. London: University of London Press Ltd.

Greenlaw, J.M.& Kurth, R.J. (1980). Current Research and Analysis of Definitions of Comprehension. Paper presented at the Annual Meeting of the American Reading Conference, Sarasota, Florida.

Helfeldt, J.P., Henk, W.A., & Fotos, A.(1986). A test of alternative cloze test formats at the sixth grade level. Journal of Educational Research, 79 (4), 216-221.

Henk, W.A.(1981). Effects of modified deletion strategies and scoring procedures on cloze test performance. Journal of Reading Behaviour, 13 (4), 346-357.

Henk, W.A.(1982). A response to Shanahan, Kamil, and Tobin: The case is not yet clozed. Reading Research Quarterly, 17 (4), 591-595.

Henk, W.A.& Selders, M.L.(1984). A test of synonymic scoring of cloze passages. The Reading Teacher, 38 (3), 282-287.

Irwin, J.W.(1991). Teaching Reading Comprehension Processes (2nd Ed.). New Jersey: Prentice Hall.

Johnson, P.H.(1983). Reading Comprehension Assessment: A Cognitive Basis. Delaware: International Reading Association.

Jonz, J. (1990). Another turn in the conversation: What does cloze measure? TESOL Quarterly, 24 (1), 61-83.

Judd, C.M., Smith, E.R., & Kidder, L.H.(1991). Reasearch Methods in Social Relations (6th Ed.). Orlando, Florida: Holt, Rinehart and Winston.

Kibby, M.W.(1980). intersentential processes in reading comprehension. Journal of Reading Behaviour, 7 (4), 299-312.

Latham, R.& Sloan, P.(1979). A Modern View of Reading. Melbourne: Nelson.

Layton, K.(1983). A Correlational Study of the Relationship between Cloze Procedures and Standardised Reading Tests for Intermediate Grades. Masters Theseis, Southwest Missouri State University.

Lewis, D.G.(1967). Statistical Methods in Education. London: University of London Press Ltd.

Lipson, M.L. & Wixson, K.K. (1991). Assessment and Instruction of Reading Disability. New York: Harper Collins.

McLeod, J.(1977). Gap Manual. Victoria: Heinemann Educational Australia.

Mossenson, L., Hill, P.& Masters, G.(1987). Torch Tests of Reading Comprehension Manual. Victoria: Australian Council for Educational Research.

Parker, R.(1985). Towards a model of the reading process. In Burnes, D. & Page, G.(Eds.). Insights and Strategies for Teaching Reading. Sydney: Brace Jovanovich. (pp.21-44).

Pitts, M.M.(1983). Comprehension monitoring: Definition and practice. Journal of Reading, 26 (6), 516-523.

Rand, E.(1978). The effects of test length and scoring method on the precision of cloze test scores. Workpapers in Teaching English as a Second Language, 7, 62-71.

Rye, J.(1982). Cloze Procedure and the Teaching of Reading. London: Heinemann Educational Books.

Rye, J.(1984). Cloze and intersentential constraint. Journal of Research in Reading, 7 (2), 113-121.

Rye, J.(1985). Are cloze items sensitive to constraints across sentences? A review. Journal of Research in Reading, 8 (2), 94-105.

Schell, L.M.(1988). Dilemmas in assessing reading comprehension. The Reading Teacher, 42 (1), 12-16.

Sciarone, A.G.& Schoorl, J.J.(1989). The cloze test; or why small isn't beautiful. Language Learning, 39 (3), 415-438.

Shanahan, T.& Kamil, M.L.(1983). The relationship of the concurrent and construct validities of cloze. Paper presented at the Annual Meeting of the National Reading Conference, Austin Texas.

Shanahan, T., Kamil, M.L.& Tobin, A.W.(1982). Cloze as a measure of intersentential comprehension. Reading Research Quarterly, 17 (2), 229-255.

Sinatra, R.& Dowd, C.A.(1991). Using syntactic and semantic clues to learn vocabulary. Journal of Reading, 35 (3), 224-229.

Slavin, R.E.(1984). Research methods in Education. New Jersey: Prentice-Hall.

Sloan, P.& Whitehead, D.(1986). Reading theories explained. Primary English Notes, 55, 1-8.

Smith, F.(1985). Reading (2nd Ed.). Cambridge: Cambridge University Press.

Soudek, M.& Soudek, L.I.(1983). Cloze after thirty years: new uses in language teaching. ELT Journal, 37 (1), 335-340.

Sternberg, R.J.(1991). Are we reading too much into reading comprehension tests? Journal of Reading, 34 (7), 540-545.

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

Taylor, W L.(1956). Recent developments in the use of cloze procedures. Journalism Quarterly, 33, 42-48.

Valmont, W.J.(1983). Cloze deletion patterns: How deletions are made makes a big difference. The Reading Teacher, 37 (2), 172-175.

Wood, N.V.(1988). Standardised reading tests and the postsecondary reading curriculum. Journal of Reading, 32 (3), 224-230.

Zakaluk, B.V. & Samuels, S.J.(Eds.).(1988). Readability: Its Past, Present, and Future. Newark, Delaware: International Reading Association.

Appendices

Appendix A

Cloze Test A1

Story - Kiya the Gull

50 Word Deletions Beginning at the Fifth Word

CLOZE TEST A1

KIYA THE GULL

by Fen. H. Lasell

Kiya was a free bird. From dawn to sundown _____ would sweep the sky, _____ sand, the sea, looking _____ food. He kept a _____ eye on the harbour _____ pick up after the _____. On the beach he _____ the children and ate _____ sandwiches they could'nt finish. _____ all sea gulls he _____ willing to eat almost _____.

Kiya began the day _____ dawn. He left the _____ island where he lived _____ went right to the _____. He wanted to be _____ when the fishermen came _____ clean their fish.

This _____ no fishermen were in, _____ the sea had left _____ the rocky beach a _____ of seaweed, crabs, and _____ - all tangled in wire.

_____ flew around the bundle. _____ he had found it _____ safe, he went to _____. He had to pick _____ the seaweed to reach _____ snails and crabs. But _____ wire kept getting in _____ way.

Kiya had to _____ at the wire until _____ end came free. Then _____ put his head in _____ opening and pulled out _____ seaweed.

Other gulls, hungry _____ breakfast, joined him. They _____ at the wire and _____ out the seaweed. They _____ what they could. They _____ away with snails in _____ bills to break the _____ shells by dropping them _____ rocks.

When the party _____ over, Kiya found that _____ it was he who _____ tangled in the wire.

"_____ - kiya - kiya," he cried.

_____ harder he tried to _____ himself, the tighter the _____ pulled. At last he freed his wings, but a loop of wire bound his back and one leg so tightly that he could not move it.

Appendix B

Cloze Test A2

Story - Kiya the Gull

50 Word Deletions Beginning at the Sixth Word

CLOZE TEST A2

KIYA THE GULL

by Fen. H. Lasell

Kiya was a free bird. From dawn to sundown he
_____ sweep the sky, the _____, the sea, looking for
_____. He kept a close _____ on the harbour to
_____ up after the fishermen. _____ the beach he
watched _____ children and ate the _____ they couldn't
finish. Like _____ sea gulls he was _____ to eat almost
anything.

_____ began the day at _____. He left the
little _____ where he lived and _____ right to the
harbour. _____ wanted to be there _____ the fishermen
came to _____ their fish.

This morning _____ fishermen were in, but
_____ sea had left on _____ rocky beach a bundle
_____ seaweed, crabs, and snails - _____ tangled in
wire.

Kiya _____ around the bundle. When _____ had
found it quite _____, he went to work. _____ had to
pick out _____ seaweed to reach the _____ and crabs.
But the _____ kept getting in his _____.

Kiya had to tug _____ the wire until one
_____ came free. Then he _____ his head in the
_____ and pulled out the _____.

Other gulls, hungry for _____, joined him. They pulled _____ the wire and pulled _____ the seaweed. They ate _____ they could. They flew _____ with snails in their _____ to break the hard _____ by dropping them on _____.

When the party was _____, Kiya found that now _____ was he who was _____ in the wire.

"Kiya - _____ - kiya," he cried.

The _____ he tried to free _____, the tighter the wire _____. At last he freed his wings, but a loop of wire bound his back and one leg so tightly that he could not move it.

Appendix C

Cloze Test A3

Story - Kiya the Gull

100 Word Deletions Beginning at the Fifth Word

CLOZE TEST A3

KIYA THE GULL

by Fen. H. Lasell

Kiya was a free bird. From dawn to sundown _____ would sweep the sky, _____ sand, the sea, looking _____ food. He kept a _____ eye on the harbour _____ pick up after the _____. On the beach he _____ the children and ate _____ sandwiches they could'nt finish. _____ all sea gulls he _____ willing to eat almost _____.

Kiya began the day _____ dawn. He left the _____ island where he lived _____ went right to the _____. He wanted to be _____ when the fishermen came _____ clean their fish.

This _____ no fishermen were in, _____ the sea had left _____ the rocky beach a _____ of seaweed, crabs, and _____ - all tangled in wire.

_____ flew around the bundle. _____ he had found it _____ safe, he went to _____. He had to pick _____ the seaweed to reach _____ snails and crabs. But _____ wire kept getting in _____ way.

Kiya had to _____ at the wire until _____ end came free. Then _____ put his head in _____ opening and pulled out _____ seaweed.

Other gulls, hungry _____ breakfast, joined him. They _____ at the wire and _____ out the seaweed. They _____ what they could. They _____ away with snails in _____ bills to break the _____ shells by dropping them _____ rocks.

When the party _____ over, Kiya found that _____ it was he who _____ tangled in the wire.

"_____ - kiya - kiya," he cried.

_____ harder he tried to _____ himself, the tighter the _____ pulled. At last he _____ his wings, but a _____ of wire bound his _____ and one leg so _____ that he could not _____ it.

A boy was _____ in his boat watching _____ gulls. When he saw _____ trouble, he got out _____ ran toward the bird._____ frightened Kiya flapped his _____ and rose out of _____, even though the wire _____ into his back and _____.

The bird glided over _____ the sandy beach and _____ a clumsy landing on _____ foot. He hopped along _____ cool, hard sand near _____ water, dragging part of _____ wire that bound him.

_____ were already gathering on _____ beach for a day _____ the sun.

"Look at _____ seagull!" someone called "He's _____ tangled up in something."

_____ ran toward Kiya. Hands _____ out for him. Beating _____ wings, Kiya managed to _____ himself again.

He flew _____ the high dune where _____ seagulls perch at noon. _____ other gulls were still _____ looking for their morning _____. Hungry as Kiya was, _____ hurt him too much _____ fly. He wanted only _____ be left in peace.

_____ children came to climb _____ dune and slide down _____ warm sand.

"Look!" one _____. "That seagull is all _____ up. Let's catch him.!"

_____ watched the children come _____ and closer. He spread _____ wings and tried to _____. At last he rose _____ of their reach.

Kiya _____ on the water. The _____ waves washed his cuts. _____ bobbed on the waves and let them draw him out to sea. How was he going to get rid of the wire?

Appendix D

Cloze Test A4

Story - Kiya the Gull

100 Word Deletions Beginning at the Sixth Word

KIYA THE GULL

by Fen. H. Lasell

Kiya was a free bird. From dawn to sundown he
_____ sweep the sky, the _____, the sea, looking for
_____. He kept a close _____ on the harbour to
_____ up after the fishermen. _____ the beach he
watched _____ children and ate the _____ they couldn't
finish. Like _____ sea gulls he was _____ to eat almost
anything.

_____ began the day at _____. He left the
little _____ where he lived and _____ right to the
harbour. _____ wanted to be there _____ the fishermen
came to _____ their fish.

This morning _____ fishermen were in,but
_____ sea had left on _____ rocky beach a bundle
_____ seaweed, crabs, and snails - _____ tangled in
wire.

Kiya _____ around the bundle. When _____ had
found it quite _____, he went to work. _____ had to
pick out _____ seaweed to reach the _____ and crabs.
But the _____ kept getting in his _____.

Kiya had to tug _____ the wire until one
_____ came free. Then he _____ his head in the
_____ and pulled out the _____.

Other gulls, hungry for _____, joined him. They pulled _____ the wire and pulled _____ the seaweed. They ate _____ they could. They flew _____ with snails in their _____ to break the hard _____ by dropping them on _____.

When the party was _____, Kiya found that now _____ was he who was _____ in the wire.

"Kiya - _____ - kiya," he cried.

The _____ he tried to free _____, the tighter the wire _____. At last he freed _____ wings, but a loop _____ wire bound his back _____ one leg so tightly _____ he could not move _____.

A boy was sitting _____ his boat watching the _____. When he saw Kiya's _____, he got out and _____ toward the bird. The _____ Kiya flapped his wings _____ rose out of reach, _____ though the wire cut _____ his back and leg.

_____ bird glided over to _____ sandy beach and made _____ clumsy landing on one _____. He hopped along the _____, hard sand near the _____, dragging part of the _____ that bound him.

People _____ already gathering on the _____ for a day in _____ sun.

"Look at the _____!" someone called "He's all _____ up in something."

People _____ toward Kiya. Hands reached _____ for him. Beating his _____, Kiya managed to raise _____ again.

He flew to _____ high dune where the _____ perch at noon. The _____ gulls were still away _____ for their morning meal. _____ as Kiya was, it _____ him too much to _____. He wanted (.y to _____ left in peace.

Then _____ came to climb the _____ and slide down the _____ sand.

"Look!" one cried. "_____ seagull is all tangled _____. Let's catch him.!"

Kiya _____ the children come closer _____ closer. He spread his _____ and tried to fly. _____ last he rose out _____ their reach.

Kiya landed _____ the water. The cool _____ washed his cuts. He _____ on the waves and let them draw him out to sea. How was he going to get rid of the wire?

Cloze Test B1

Story - Hattie the Backstage Bat

50 word Deletions Beginning at the Fifth Word

CLOZE TEST B1

Hattie, the Backstage Bat

by Don Freeman

The backstage of a dark, empty theatre is a lonely place where only a bat would feel at home. To a lttle bat _____ Hattie, this was home.

_____ had lived in the _____ Theatre all her life, _____ she had never seen _____ green tree or a _____ house. She had never _____ in the bright moonlight _____ way other bats do.

_____ Hattie, the sky was _____ space high above the _____. Every night she flew _____ for hours at a _____, sweeping in and out _____ rafters and between the _____ curtains.

Then, when she _____ tired, she landed on _____ rope, folded her wings _____ against her sides, and _____ upside down by her _____ to sleep.

The only _____ who knew about Hattie _____ Mr. Collins. He came _____ every morning to clean.

_____ hadn't been a show _____ the old Grand Theater _____ quite a while, but _____ Collins was never lonely. _____ had Hattie. Once he _____ her a little hat _____ of things he found _____ an old trunk.

Each _____ Mr. Collins pulled his _____ to the middle of _____ stage and shared his _____ with his friend. He _____ that bats liked to _____ flowers, so he always _____ Hattie daisies.

106

While they _____, Hattie listened as Mr. _____

talked about the wonderful _____ that had been given

_____ this very stage.

One _____ he had very important _____ to tell

Hattie. "Starting _____, some actors are coming _____ to

rehearse a new _____," he said. "That means _____ will

have to stay _____ of sight. I don't _____ why, but

people get very frightened when they see a bat flying around."

Then Mr. Collins began to shoo Hattie into the rafters.

Appendix F

Cloze Test B2

Story - Hattie the Backstage Bat

50 Word Deletions Beginning at the Sixth Word

CLOZE TEST 32

Hattie, the Backstage Bat

by Don Freeman

The backstage of a dark, empty theatre is a lonely place where only a bat would feel at home. To a lttle bat named _____, this was home.

Hattie _____ lived in the Grand _____ all her life, so _____ had never seen a _____ tree or a haunted _____. She had never flown _____ the bright moonlight the _____ other bats do.

For _____, the sky was the _____ high above the stage. _____ night she flew about _____ hours at a time, _____ in and out of _____ and between the stage _____.

Then, when she was _____, she landed on a _____, folded her wings tightly _____ her sides, and hung _____ down by her claws _____ sleep.

The only one _____ knew about Hattie was _____ Collins. He came in _____ morning to clean.

There _____ been a show in _____ old Grand Theater for _____ a while, but Mr. _____ was never lonely.He _____ Hattie. Once he made _____ a little hat out _____ things he found in _____ old trunk.

Each noontime _____ Collins pulled his chair _____ the middle of the _____ and shared his lunch _____ his friend. He knew _____ bats liked to eat _____, so he always brought _____ daisies. '

109

While they ate, _____ listened as Mr. Collins _____ about the wonderful plays _____ had been given on _____ very stage.

One afternoon _____ had very important news _____ tell Hattie. "Starting today, _____ actors are coming here _____ rehearse a new play," _____ said. "That means you _____ have to stay out _____ sight. I don't know _____, but people get very frightened when they see a bat flying around."

Then Mr. Collins began to shoo Hattie into the rafters.

Appendix G

Cloze Test B3

Story - Hattie the Backstage Bat

100 Word Deletions Beginning at the Fifth Word

Hattie, the Backstage Bat

by Don Freeman

The backstage of a dark, empty theatre is a lonely place where only a bat would feel at home. To a lttle bat _____ Hattie, this was home.

_____ had lived in the _____ Theatre all her life, _____ she had never seen _____ green tree or a _____ house. She had never _____ in the bright moonlight _____ way other bats do.

_____ Hattie, the sky was _____ space high above the _____. Every night she flew _____ for hours at a _____, sweeping in and out _____ rafters and between the _____ curtains.

Then, when she _____ tired, she landed on _____ rope, folded her wings _____ against her sides, and _____ upside down by her _____ to sleep.

The only _____ who knew about Hattie _____ Mr. Collins. He came _____ every morning to clean.

_____ hadn't been a show _____ the old Grand Theater _____ quite a while, but _____ Collins was never lonely. _____ had Hattie. Once he _____ her a little hat _____ of things he found _____ an old trunk.

Each _____ Mr. Collins pulled his _____ to the middle of _____ stage and shared his _____ with his friend. He _____ that bats liked to _____ flowers, so he always _____ Hattie daisies.

While they _____, Hattie listened as Mr. _____ talked about the wonderful _____ that had been given _____ this very stage.

112

One _____ he had very important _____ to tell Hattie. "Starting _____, some actors are coming _____ to rehearse a new _____," he said. "That means _____ will have to stay _____ of sight. I don't _____ why, but people get _____ frightened when they see _____ bat flying around."

Then _____ Collins began to shoo _____ into the rafters. "I'm _____ to have to do _____, my dear," he shouted, "_____ it's for your own _____ as well as mine. _____ want the play to _____ a success, don't we?"

_____ Hattie did as she _____ told. The actors had _____ idea a bat hung _____ above them as they _____ reading their parts in _____ play.

Day after day, _____ actors came in to _____ their parts until they _____ all their lines by _____. Day after day, Hattie _____ well out of sight. _____ was only late at _____ that Hattie flew down _____ the stage and ate _____ food that Mr. Collins _____ left for her.

One _____ Hattie heard the sound _____ hammering. The set for _____ play was being built _____ the stage. Hattie saw, _____ the first time in _____ life, not only a _____ but a three storey _____ It was a haunted _____. It even had a _____ attic made to order _____ a bat.

Weeks went _____, and finally there was _____ dress rehearsal. Hattie watched _____ surprise as an actor, _____ a long black cape _____ looking like a huge _____, began to climb in _____ out of the windows _____ the house.

113

"Why does'nt _____ fly the way I _____?" she asked herself. "I _____ show that actor how _____ act like a bat." _____ Hattie didn't dare move from the rafters. Then, at last, it was opening night.

Appendix H

Cloze Test B4

Story - Hattie the Backstage Bat

100 Word Deletions Beginning at the Sixth word.

CLOZE TEST B4

Hattie, the Backstage Bat

by Don Freeman

The backstage of a dark, empty theatre is a lonely place where only a bat would feel at home. To a lttle bat named _____, this was home.

Hattie _____ lived in the Grand _____ all her life, so _____ had never seen a _____ tree or a haunted _____. She had never flown _____ the bright moonlight the _____ other bats do.

For _____, the sky was the _____ high above the stage. _____ night she flew about _____ hours at a time, _____ in and out of _____ and between the stage _____.

Then, when she was _____, she landed on a _____, folded her wings tightly _____ her sides, and hung _____ down by her claws _____ sleep.

The only one _____ knew about Hattie was _____ Collins. He came in _____ morning to clean.

There _____ been a show in _____ old Grand Theater for _____ a while, but Mr. _____ was never lonely.He _____ Hattie. Once he made _____ a little hat out _____ things he found in _____ old trunk.

Each noontime _____ Collins pulled his chair _____ the middle of the _____ and shared his lunch _____ his friend. He knew _____ bats liked to eat _____, so he always brought _____ daisies.

While they ate, _____ listened as Mr.Collins _____ about the wonderful plays _____ had been given on _____ very stage.

One afternoon _____ had very important news _____ tell Hattie. "Starting today,_____ actors are coming here _____ rehearse a new play," _____ said. "That means you _____ have to stay out _____ sight. I don't know _____, but people get very _____ when they see a _____ flying around."

Then Mr. _____ began to shoo Hattie _____ the rafters. "I'm sorry _____ have to do this, _____ dear," he shouted, "but _____ for your own good _____ well as mine. We _____ the play to be _____ success, don't we?"

So _____ did as she was _____. The actors had no _____ a bat hung high _____ them as they sat _____ their parts in the _____.

Day after day, the _____ came in to rehearse _____ parts until they knew _____ their lines by heart. _____ after day, Hattie kept _____ out of sight. It _____ only late at night _____ Hattie flew down to _____stage and ate the _____ that Mr. Collins had _____ for her.

One morning _____ heard the sound of _____. The set for the _____ was being built on _____ stage. Hattie saw, for _____ first time in her _____, not only a tree _____ a three storey house. _____ was a haunted house. _____ even had a tower _____ made to order for _____ bat.

117

Weeks went by, _____ finally there was a _____ rehearsal. Hattie watched in _____ as an actor, wearing _____ long black cape and _____ like a huge bat, _____ to climb in and _____ of the windows of _____ house. .

"Why does'nt he _____ the way I do?" _____ asked herself. "I could _____ that actor how to _____ like a bat." But _____ didn't dare move from the rafters.

Then, at last, it was opening night.

Appendix J

Cloze Test Scoring Keys

Cloze Tests A1 and A3

Story - Kiya the Gull

Scoring Key

The first word is the original word from the passage.

1) he - Kiya

2) the - search

3) for

4) close - watchful, keen, sharp

5) to

6) fishermen - boats, people, humans
   sailors

7) watched - saw, spotted, followed

8) the - leftover, some

9) like - of

10) was - is

11) anything - everything

12) at - before

13) little - deserted, small, beautiful,
    tiny, sandy, big

14) and - then

15) harbour - mainland, beach, jetty,
    port, shore

16) there - near, first, around,
    early, close, ready

17) to

18) morning - day, time

19) but - and, because, for, however, yet

20) on - along, at

21) bundle - bed, row,
    smorgasboard, mass,
    clump

22) snails

23) Kiya - he

24) when - once, after

25) quite - was, very

26) work - eat, land, look,
    investigate, scavenge

27) out - up, at, through,
    off

28) the - for

29) the

30) his - the

31) tug - pull, pick, peck,
    nibble, bite

32) one - the, an, it's

33) he - kiya

34) the -an

35) the -some

36) for - at

120

37) pulled - tugged, pecked

went, picked, nibbled, bit,

tore

38) pulled - picked, jerked,

dragged, cleared, took,

dug

39) ate - did

40) flew - went

41) their

42) hard - snail, brittle,

small

43) on - against

44) was - had

45) now - then, unfortunately,

surprisingly, unhappily,

amazingly

46) was

47) Kiya

48) the - but

49) free - untangle, lift, save,

release

50) wire

51) freed - flapped, fanned, tried,

spread, released, untangled.

52) loop - piece, strand, coil, bit,

thread, length

53) back - beak, body, foot, neck,

tail, wing, feathers

54) tightly

55) move - straighten,

free, use, release,

loosen, undo

56) sitting - out, nearby,

sailing, drifting,

rowing, standing, lying

57) the - sea, for

58) Kiya's - the, this

59) and

60) the - very, being, a

61) wings

62) reach - trouble,

sight, danger

63) cut - dug, wound, tore,

pressed, stuck, stabbed

64) leg - legs, wings,

neck

65) to - towards, past,

above, onto

66) made - performed, did

67) one - his

68) the - over, to

69) the

70) the

71) people - families,

fisherman, children

72) the

73) in

74) the - that

75) all - got, badly, tightly

76) people - they, children,
   he, everyone, someone

77) reached - stretched, held,
   reaching, grabbed

78) his

79) raise - free, save, lift,
   move

80) to - above, over, towards,
   onto, around

81) the - many, some, all,
   other

82) the - some

83) away - out, there, noisily,
   hungrily

84) meal - food, breakfast,
   feed, snack

85) it - flying

86) to

87) to

88) then - some, the, many

89) the - over

90) the

91) cried - yelled, shouted,
   said, called, exclaimed

92) tangled - caught,
   tied

93) Kiya - he

94) closer

95) his

96) fly - escape

97) out

98) landed - floated,
   settled

99) cool - gentle, ocean,
   salty, small, smooth,
   rough, big, huge

100) he - Kiya

Cloze Tests A2 and A4

Story - Kiya the Gull

Scoring Key

The first word is the original word from the passage.

1) would

2) sand - ground, land

3) food - fish

4) eye - watch

5) pick - clean

6) on - at, from

7) the

8) sandwiches - food

9) all - other, most

10) willing - going, able,

    likely

11) Kiya - he

12) dawn

13) island - beach, tree,

    place

14) went - flew

15) He - Kiya

16) when - because, before

17) clean - get, catch

18) no

19) the

20) the - a

21) of

22) all

23) flew - went, pecked,

    walked, looked, circled

24) he

25) safe - interesting

26) he

27) the

28) snails

29) wire - seaweed

30) way - beak

31) at - on

32) end - piece

33) put

34) opening - wire,

    bundle, seaweed

35) seaweed - food

36) breakfast - food,

    crabs

37) at - out

38) out - at

39) what - all

40) away - off

41) bills - beaks

42) shells

43) rocks - concrete

44) over - finished

45) it

46) tangled - caught

47) Kiya - help

48) harder - more

49) himself

50) pulled - got, tangled,

became

51) his

52) of

53) and - with, to

54) that

55) it - away, anywhere,

56) in - on

57) gulls - action, bird

58) trouble

59) ran - went, swam,

walked

60) frightened - scared,

helpless

61) and

62) even

63) into - all, through

64) the

65) the - a

66) a

67) foot - leg

68) cool

69) water - rocks, sea

70) wire

71) were

72) beach - shore

73) the

74) seagull - bird, gull

75) tangled - tied, caught

76) ran - walked, came

77) out

78) wings

79) himself - up

80) the - a

81) seagulls - birds

82) other - sea

83) looking - searching,

84) hungry

85) hurt

86) fly

87) be

88) children - people

89) dune

90) warm - slippery, hot

91) that - a, the

92) up

93) watched - saw

94) and

95) wings

96) at

97) of

98) on - in

99) waves - water

100) bobbed - sat

Cloze Tests B1 and B3

Story - Hattie the Backstage Bat

Scoring Key

The first word is the original word from the passage.

1) named - called
2) Hattie - she
3) grand - dark, old, empty
4) so - and
5) a
6) haunted - normal
7) flown
8) the
9) for - to
10) the
11) stage - theatre
12) about - around
13) time
14) of
15) stage - huge, theatre, big
16) was - got, grew, had
17) a - the
18) tightly - up, in, firm
19) hung - turned, swung
20) claws
21) one - person
22) was
23) in - nearly
24) there

25) in - at
26) for - in
27) Mister
28) he
29) made - got
30) out - made
31) in
32) noontime - day
33) chair
34) the
35) lunch
36) knew - thought
37) eat - smell
38) brought - gave
39) ate
40) Collins
41) plays - shows
42) on
43) afternoon - day
44) news
45) today - tomorrow
46) here - over, in
47) play - show
48) you

125

49) out

50) know

51) very - really, quite

52) a

53) Mister

54) Hattie - her

55) sorry

56) this - it

57) but

58) good

59) we

60) be

61) so

62) was

63) no

64) high - directly, up,

   right, way

65) sat - were

66) the

67) the

68) rehearse - practice

69) knew - learned

70) heart

71) kept - hid, stayed

72) it

73) night

74) to - onto

75) the - some

76) had

77) morning - evening, day

78) of

79) the

80) on - behind

81) for

82) her

83) tree - house

84) house

85) house

86) tower - big, small

87) for

88) by

89) a

90) in - with

91) wearing - in, with

92) and

93) bat

94) and

95) of - in

96) he

97) do - can, fly

98) could - will, can,

   should

99) to

100) but - still

Cloze Tests B2 and B4

Story - Hattie the Backstage Bat

Scoring Key

The first word is the original word from the passage.

1) Hattie

2) had

3) theatre

4) she - Hattie

5) green - live, lovely,
   big, nice, real

6) house

7) in - through, into, by

8) way

9) Hattie - her

10) space - roof, curtain

11) every - at, that, one

12) for

13) sweeeping - flying,
    weaving, going

14) rafters - chairs, doors,
    rooms, posts

15) curtains - seats, lights,
    chairs, objects

16) tired - sleepy

17) rope - seat, chair, rafter,
    board, perch, bar

18) against - on, beside, underneath,
    by, around

19) upside

20) to

21) who - that

22) Mister

23) every - the, each

24) hadn't

25) the

26) quite

27) Collins

28) had - loved, liked

29) her - Hattie

30) of

31) an - the, his

32) Mister

33) to - from, onto, in

34) stage - theatre, room

35) with

36) that

37) flowers - daisies,
    food

38) Hattie - some

39) Hattie - she

40) talked - told, spoke

41) that

42) this - that

43) he

44) to

45) some - the

46) to

47) he

48) will - Hattie

49) of

50) why - where, them

51) frightened - scared, angry,

   mad, nervous

52) bat

53) Collins

54) into

55) to - I

56) my - Hattie

57) it's

58) as

59) want

60) a

61) Hattie - she

62) told

63) idea

64) above

65) reading - rehearsing

66) play - theatre

67) actors

68) their

69) all

70) day

71) well - in, herself, way

72) was

73) that - when

74) the

75) food - scraps, daises,

   flowers

76) left - brought, made,

   picked

77) Hattie - she

78) hammering - people,

   builders, hammers

79) play

80) the

81) the

82) life

83) but

84) it

85) it

86) attic - especially,

   just, they

87) a - the

88) and

89) dress - last, final

90) surprise - awe,

   amazement

91) a

92) looking - ears

93) began

94) out

95) the - a

96) fly - move

97) she - Hattie

98) show - teach

99) act - fly, move

100) Hattie - she