VALIDATION STUDY OF THE HICKS CLOZE-READING TEST FOR GRADES TWO - SIX

By

DONNA WHEELER HICKS

Bachelor of Science in Education
Central State University
Edmond, Oklahoma
1958

Master of Teaching
Central State University
Edmond, Oklahoma
1961

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF EDUCATION
May, 1974

VALIDATION STUDY OF THE HICKS CLOZE-READING TEST FOR GRADES TWO - SIX

Thesis Approved:

Daniel D. Ray
Thesis Adviser

Dean of the Graduate College

ACKNOWLEDGMENTS

The writer wishes to express her sincere appreciation and gratitude to Dr. Darrel D. Ray, chairman of the doctoral committee, for his patient supervision, encouragement, and many constructive criticisms during the preparation and development of this study. For their helpful suggestions, the writer also wishes to express appreciation to the other members of her committee, Dr. Josephine Hoffer, Dr. Bill Elsom, Dr. Russell Dobson, and Dr. William Segall.

Appreciation is extended to Mr. Ronald Bowers, the faculty and students from the Chandler Eastside School and to Mr. Mike Rafferty, the faculty and students from the Perkins Elementary School. The courtesy and cooperation extended by these administrators, teachers, and students greatly facilitated the testing program.

A special word of thanks is extended to my colleagues and friends for their assistance with the testing program. Appreciation is also expressed to my colleague Miss Janette Staton for her help and encouragement, to Dr. David Elijah for his assistance with the statistical part of the investigation, and to Mrs. Linda Baker for her help in typing the final paper.

Without the love and support of my husband, Raymond, this study could not have been completed. To my daughters, Essa and Paula, this study is dedicated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

THE NATURE OF THE PROBLEM

Introduction

This study was concerned with the validation of the Hicks Cloze-Reading Test, an instrument to be used by the classroom teacher in assessing the reading ability level of children. There has been continued research through the years toward the development of more precise measurements of reading ability. Some aspects of the evolving research have been concerned with the cloze procedure and the implications it has for educators concerned with the reading process.

General Background of the Study

The cloze procedure is not a new concept in education, having been developed by Wilson Taylor (1953) as a way of measuring the use of language. Taylor proposed two rationales for the procedure. One was based on the Gestalt theory of closure and the other was based on an information or communication theory as developed by Miller (1951).

The word "cloze" was coined by Taylor from the Gestalt concept of "closure," which is a tendency for an individual to complete a familiar pattern by mentally closing the gaps in a structure; to see a "broken circle as a whole one." The individual can complete the broken circle even with the missing parts because the circle is a familiar pattern. A "cloze unit" as defined by Taylor is:

> . . . any single occurence of a single attempt to reproduce accurately a part deleted from a 'message' [any language product] by deciding, from the context that remains, what the missing part should be (1953, page 416).

Using this frame of reference, Taylor (1953) describes the logic of the cloze procedure as a method of:

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

Because the cloze unit measures language-usage correspondence instead of meaning, it is classified as a common denominator of communication success.

Taylor (1953) states that "the notion of cloze procedure was 'sparked' by implications of Osgood's learning theory of communication" (page 418). Redundancies, transitional probabilities of language to the development of dispositional mechanisms--all play a part in the transmission and receiving of messages. Redundancy, which in information theory means "predictability," is defined as being an "estimate of how effectively the factors in a situation combine to restrict the results of that situation toward a single kind of outcome (event)." The converse of redundancy is "entropy" or "an estimate of the lack of organization or the amount of uncertainty in a specified 'system' or set of events" (Taylor, 1954, page 15).

Osgood (1952) explains redundancy as a repetition of meaning, as an internal tie between words, which makes it possible to replace the "little" words. An example of this: "Man coming" means the same as "A man is coming this way now." In the latter sentence the singular number of the subject is mentioned three times, the present tense twice,

and the action of the direction twice. Several of the "little" words could be removed from the sentence without losing the meaning of the sentence.

In learning to "think in" a language, an individual develops a number of verbal skill patterns which tend to become automatic. These habits reflect the redundancies and transitional probabilities of the language patterns and skills involved. As a result of these events, each individual develops his own set of habits. If one set of habits corresponds with the set of another individual, communication can occur. When words come in sequences that best fit the existing receiving habits of a reader, the reader perceived with more clarity the intended meaning.

These terms are explained by Osgood (1952) as part of the habits of expression of an individual. Habits or reading or listening cause the individual to anticipate words, almost automatically, when he is receiving messages. Often, upon seeing a phrase that looks familiar, the individual will immediately complete it in his own way even though the written phrase may end differently.

The concept of cloze procedure involves both oral and written communication and does not inherently require any particular kind of deletion. The cloze procedure measures the effects that the elements of language (sentences, words, and syllables) may have on the readability of a particular passage. Taylor (1953) states that the cloze method seems to deal with more-or-less parallel sets of meaning-pattern relationships. Different persons may express the same meaning in somewhat differing ways, and the same language patterns may have differing meanings for different people. The cloze procedure, then, takes a measure

of the likeness between the patterns a writer has used and the patterns the reader is anticipating while he is reading. To the extent that the reader and the writer have similar experience backgrounds, interests, and language habits, the reader should be able to make accurate predictions of the words which have been deleted.

Miller (1951) discusses language patterns and the communication process when he states that:

> Every communication must have a source and a destination for the information that is transferred, and these must be distinct in space or time. Between the source and the destination there must be some link that spans the intervening space or time, and this link is called a communication channel. In order that the information can pass over the channel, it is necessary to operate on it in such a way that it is suitable for transmission, and the component that performs this operation is a transmitter. At the destination there must be a receiver that converts the transmitted information into its original form. These five components--source, transmitter, channel, receiver, and destination--comprise the idealized communication system. In one form or another, these five components are present in every kind of communication (page 6).

Closure is part of a decoding process and as such becomes part of a restructuring or "recoding" operation. Miller (1960) theorizes that the operation of the transmitter in writing the passage is often referred to as "encoding." The code is the pattern of energies that can travel over the connecting link. The receiver reverses the operation of the transmitter and reconverts the coded message into a more usable form. Thus, the operation of the receiver is referred to as "decoding" or the interpretation of the passage.

Weaver (1965) explains this decoding process as a planned sequence of words. The cloze procedure, however, interrupts this decoding process requiring the reader to search for a direct cue to the language element which should occupy the space. The reader at one point is

reading; at another point the reader must produce a word to fit a certain context utilizing both decoding and encoding skills. The cloze procedure, therefore, "enlists the subject in a hierarchical process which goes beyond the ordinary demands of reading" (page 177).

In utilizing the cloze procedure, a cloze test can be constructed through a systematic deletion of words, substituting a blank line of a predetermined length in the same space. Subjects taking the test are instructed to predict from the remaining context what word belongs in each space. There are two basic types of cloze deletions: structural, in which every nth word in a passage is deleted; and lexical, in which nth noun or main verb is deleted. Rankin (1957) is credited with adapting the cloze procedure to the division of language into the structural (an interrelationship between ideas) and lexical (substantive content of a message) elements. The structural deletion is often confounded with the lexical deletion, however, because of the method used in forming the structural cloze.

Cloze procedure is something like a sentence-completion test in that the subject is presented with incomplete sentences and there are blanks to be filled in from context. Since this form of test measures specific knowledge, the items to be deleted must be selected accordingly. Cloze does not have isolated answers but utilizes a contextually interrelated series of blanks. The cloze does not deal directly with specific meaning but is used to compare the extent of the similarity of the language patterns used by a writer to express a meaning intended for the reader.

Cloze research has determined the procedure to be a valid one for measuring general reading achievement. Substantial correlations have

been noted between cloze tests and various standardized reading tests even though the cloze tests were based on a variety of materials and were constructed and administered in different ways. Rankin (1959) noted that the cloze procedure has a sufficiently high correlation with various validity criteria signifying that it has considerable concurrent validity and concluded that there is ample evidence upholding "the validity and usefulness of this technique as a measuring instrument."

Weaver and Kingston (1963) conducted a factor analytic study of the relationship of two cloze procedures to standardized reading tests. Both structural and lexical deletions were utilized as part of the eight cloze tests, a portion of a battery involving 18 cognitive tests which were then submitted to a factor analysis. Three factors were found: 1) verbal comprehension, 2) redundancy utilization, and 3) rote memory. The cloze tests were found to be most related to "redundancy utilization," which is similar to Taylor's predictability or restrictive factor.

Although this study involved the implementation of the cloze procedure as a measurement device, extensive research is available in the literature pertaining to the use of the cloze procedure in other areas. Major areas of focus include: 1) cloze as a quantitative measure of readability, 2) cloze as a tool for investigating language variables, and 3) cloze as a suitable teaching device. The Hicks Cloze-Reading Test, as designed, is a tool for measuring the reading ability levels of students at both the primary and intermediate levels.

Need for the Study

Adequately teaching a classroom of children is a challenge facing many educators today. This challenge is compounded by the many ability ranges, interests, experiences, and motivations found within any given classroom.

Goodlad and Anderson (1963) present several generalizations about pupil ability levels with which the elementary school teachers must deal as each new year begins and progresses. Children enter the first grade with a range of from three to four years in their readiness to learn. This initial spread in abilities increases over the years and has approximately doubled by the end of the elementary school years.

Teachers are desirous of finding a means of narrowing the range of reading levels within the classroom. This becomes a real challenge as the teacher begins to identify and assign to students a variety of suitable reading materials that will be both interesting and informative, but not frustrating.

Procedures for determining the reading levels vary among schools and among teachers. Most often standardized achievement tests or standardized diagnostic reading tests, administered individually or to the entire class, are used to determine instructional levels within a classroom. The reasons for their popularity can be readily seen: they are easily administered and easily scored.

Standardized achievement tests have been shown to overestimate the reading ability of youngsters, often placing them in a frustrating learning experience (Killgallon, 1942; Botel, 1957; Sipay, 1961; McCracken, 1962; Millsap, 1963). Botel found at the second year level,

only 11 percent of the youngsters were properly rated, while 85 percent were overrated and 4 percent were underrated. On the intermediate level 33 percent were rated properly, while 33 percent were overrated and 33 percent were underrated. Betts (1957) claims that the standardized survey tests frequently rate children from 1 to 4 grades above their actual achievement level. The research conducted by Sipay (1961) with fourth grade pupils supports this claim. The comparison was between standardized reading tests and two informal reading inventories. Sipay's conclusions were that when using a more stringent criterion in the interpretation of the formal instrument, the Metropolitan Reading Test overestimated the instructional level by .79 of a grade level, while the Gates Reading Survey and the California Reading Test scored .97 and 1.7 grade levels respectively above the informal instructional level. McCracken (1962) concluded that the standardized survey test scores would place 63 percent of the subjects at the frustration level and 93 percent at a book level too difficult for instructional comfort if these scores were used to determine book placement. McCracken further stated that the use of standardized reading tests necessitates that considerable caution should be exercised if the results are to be utilized to establish reading levels or instructional groups. Millsap (1963) found that teachers cannot always tell when materials are frustrating to children, with inappropriate reading materials being used in the classrooms about one-third of the time.

Another method most often used by the teacher to determine reading abilities is the informal reading inventory. This method of evaluation is accepted as being valid and reliable since it is similar to the techniques often used in classroom instruction. The informal inventory

identifies the child's independent, instructional, and frustration reading levels and, also, observes reading characteristics to which standardized tests are not sensitized (Alexander, 1968). The inventories, although relatively simple to administer, are time consuming since individual administration is necessary. McCracken (1963) established the validity of the Standard Reading Inventory, which is similar to many other informal reading inventories. This evaluative procedure is a more accurate measure than the reading achievement tests; however, interpretation is dependent upon the skills and biases of the teacher using the instrument. In view of the limitations of the standardized reading achievement tests toward the over-estimation of the reading instructional levels of youngsters and the time involved in administering the individualized reading tests, a method or technique which teachers can utilize to determine the appropriate reading levels of youngsters, that is less time consuming and more accurate, is needed.

The cloze procedure would appear to combine the advantages of both standardized and informal test procedures--the reliability, validity, and much of the scoring ease of the former, with the pertinence and the relevance of the latter (Pennock, 1973). A cloze test is easily administered in a short period of time and can be used successfully with a group. This type of informal group testing would support and reinforce the decisions of the teacher in deciding the proper levels of reading instruction for each child. Many classroom organizational and instructional problems would be eliminated. Reading instruction, with each child at the appropriate instructional level, could proceed from the first week of the school year. It is for this purpose, the development of a new evaluative reading instrument, that this study has been conducted.

Statement of the Problem

The problem of this study is to determine if there is a relationship between the reading scores on the Standard Reading Inventory and the Hicks Cloze-Reading Test at the independent, instructional, and frustration levels. To facilitate in the analysis of this problem, the investigation centered around three major questions.

1. Is there a relationship between the grade equivalency of the Hicks Cloze-Reading Test and the Standard Reading Inventory independent reading level at the second, third, fourth, fifth, and sixth grade levels?

2. Is there a relationship between the grade equivalency of the Hicks Cloze-Reading Test and the Standard Reading Inventory instructional reading level at the second, third, fourth, fifth, and sixth grade levels?

3. Is there a relationship between the grade equivalency of the Hicks Cloze-Reading Test and the Standard Reading Inventory frustration reading level at the second, third, fourth, fifth, and sixth grade levels?

Purpose of the Study

The purpose of this study was to validate the Hicks Cloze-Reading Test instrument.

Definition of Terms

Cloze test - specific application of the "cloze procedure." This procedure was developed by Taylor (1953). When a passage is mutilated through the deletion of every "nth" word, replacement by the decoder is predicated on his ability to interpret meanings of words, note grammatical relationships, and sense the encoder's writing style. The greater the number of successful cloze responses the higher the decoder's comprehension of the passage.

Cloze procedure - a random deletion of a portion of words in a passage, the replacement of deleted words with a blank of uniform length, with instructions for the subject to write in the word that best fits the context of that passage. Words are not deleted in the first or last line of a passage.

Cloze item - refers to the word deleted in a reading passage.

Raw-Cloze score - refers to the total acceptable cloze responses made by the reader on each cloze test.

Cloze procedure percentage score - refers to the relationship of acceptable cloze responses to the total deletions possible on a given reading passage. The result is recorded as a percentage score and always has a value less than 100 percent.

Readability - is the total of all those elements within a given piece of printed material that affect the success of the reader. The success is the extent to which the reader understands the passage, reads the passage at an optimum speed, and finds the passage interesting (Dale and Chall, 1948). Readability or difficulty of the material is determined by applying an accepted formula to samples of the reading materials. The Spache Readability Formula (1953) and the Dale-Chall Readability Formula (1948) will be utilized to determine the readability levels of each passage in the Hicks Cloze-Reading Test.

Hicks Cloze-Reading Test - is designed to give an independent, instructional, and frustration reading level. Passages are approximately two hundred words in length. Readability levels have been established through the use of the Spache and Dale-Chall Readability Formulas. There are two forms of this instrument, Form I and Form II. Each form

is scored for exact word deletions. Directions are provided for the child on his copy and for the teacher on the scoring copy.

Assumptions of the Study

The writer makes the following assumptions: that various readability formulas may be used to accurately determine the reading level of a passage; that new criteria can be established that will estimate reading levels from the raw-cloze test scores obtained on an every tenth word deletion.

Limitations of the Study

Although every effort was exerted to make this study as scientific as possible, it is essential to recognize its limitations. Important variables not held constant were the physical being and the emotional adjustment of pupils, the time of day during which test instruments were administered, and the environmental aspects of the testing situation. A further restriction was imposed by the measuring instrument, the Hicks Cloze-Reading Test. The validity and reliability established were dependent upon the way the instrument was developed and administered.

Because of these uncontrolled factors, it should be apparent that the findings and conclusions of this investigation can be generalized beyond the population from which this sample was drawn, or to any other instruments designed along the same lines as those employed in the study only if the limitations are fully recognized.

CHAPTER II

REVIEW OF SELECTED LITERATURE

Research utilizing the cloze procedure as a measurement of reading comprehension will be presented in this review. Taylor (1953) introduced the concept of "cloze procedure" as a new tool for measuring the effectiveness of communication with three pilot studies utilizing cloze as a measure of readability. In each study results were achieved determining cloze to be a successful measure of readability of materials. Taylor (1954) next investigated the relationship between "cloze procedure" and the entropy or "uncertainty" measure of information theory. Two prose passages were mutilated systematically to develop five deletion versions which were then administered to 287 freshman rhetoric students. Analysis of the passages yielded positive information about the utilization of contextual constraints in understanding printed material. Additional research conducted by Taylor (1957, 1958) has given much direction toward the use and application of cloze procedure.

An evaluation of the validity, performance, and utility of cloze tests was conducted by Rankin (1958). He postulated that comprehension tests emphasizing the measurement of either the substantive content of a message or the interrelationships between ideas could be constructed by varying the type of words deleted: the restriction of word deletion to nouns and verbs measures primarily the comprehension of substantive content. The postulate was confirmed. Correlations ranging from .29

to .48 were obtained between the cloze tests and the Diagnostic Reading Tests: Survey Section for the participants in a college reading improvement program. Validity coefficients of .59 and .56 were obtained for the tests as measures of pre- and post-reading knowledge of specific science materials. Rankin concluded that this form of the cloze procedure produces tests which are not very accurate measures of general reading skill but are sufficiently accurate for measuring pre- and post-reading knowledge and specific reading comprehension.

Bormuth (1962) investigated three aspects of cloze tests: 1) their validities as measures of comprehension ability, 2) their efficiencies when used for the purpose of discriminating among the difficulties of tests and the abilities of individuals, and 3) their validities as measures of the amounts of comprehension with which passages are read. Nine cloze tests covering seven comprehension skills were administered to children in the intermediate grades. The findings supported the hypothesis that cloze tests are valid measures of comprehension ability with correlations between the totals of the comprehension and cloze scores reaching .95. Significant results were also established for the other two questions investigated. Other research by Bormuth (1965, 1967) established much of the cloze criteria being used in current research, i.e., scoring methods, deletion procedures. Scoring criteria, established on the basis of comparable scores between a cloze test and multiple-choice comprehension test, have been used in many of the cloze studies. Bormuth (1968) established 57 percent cloze score as equal to the independent reading level, 44 percent as equal to the instructional reading level, and 38 percent equal to the frustration reading level.

These criteria vary from the criteria established earlier by Bormuth (1967) when 38 percent accuracy was determined to be the instructional level and 50 percent accuracy was considered as the independent reading level.

Rankin and Culhane (1969) replicated the earlier studies by Bormuth (1967, 1968) based on the two cloze scoring criteria. Five classes with 105 students were participants in the study. Each subject was: 1) administered a cloze test, 2) instructed to read the unmutilated copy, and 3) then asked to complete a multiple-choice comprehension test. Pearson correlations ranging from .54 to .77, with an average of .68, were computed between the cloze test results and the corresponding objective comprehension test results.

A scoring criterion of 50 percent for independent reading, 30 percent for instructional reading, and 20 percent for frustration reading was established by Ransom (1967) as part of a study to determine reading levels utilizing the cloze procedure. Six classes (178 pupils), representing grades one through six, participated in the study. These classes were selected by the principal to include experienced teachers and those willing to help in the study. A second study (Kirby, 1967) was conducted concurrently with the same subjects to determine relationships between the reading levels as indicated by cloze test scores and scores received on oral and silent standardized reading tests. Variables considered in the study by both examiners were sex, reading achievement, ability level, and grade level. To determine if the mental ability of the sample was representative of the total population, the Lorge-Thorndike Intelligence Tests were given. Socio-economic levels were also determined for the subjects through application of the

Minnesota Scale for Paternal Occupations. The I. Q. scores obtained ranged from 69-135 with a mean I. Q. of 101.7 and a Standard Deviation of 13.3. The median I. Q. for communities specified as average in terms of socio-economic level was comparable indicating that the sample could be compared to the total population.

Instruments used in addition to the cloze test for comparison were the Gates Reading Tests (Primary, Advanced Primary, and Survey), the Gilmore Oral Reading Tests, and the Gray Oral Reading Test. The Gates Reading Tests were used to assess the silent reading achievement of the pupils. The Gilmore Oral Reading Tests were administered individually to each pupil in order to appraise the oral reading ability in addition to yielding word accuracy and comprehension scores. The Gray Oral Reading Test was used to assess oral reading skills and to aid in diagnosing reading difficulty.

Cloze tests were constructed by the research team composed of Ransom, Kirby, and one other member. The 11 passages were 23 to 212 words in length and ranged in reading difficulty from the 1.4 to 9.8 levels. Every fifth word was deleted, making a total of 224 deletions in the whole test. The cloze test was administered to an entire classroom at one setting. Exact words were counted for a total raw score for each subject. The instructional level was determined if the subject attained 30 to 49 percent accuracy. Ransom also scored the passages in such a way to determine a raw average score. All of the scores were added and then the total was divided by the number of passages attempted by the subject. The 50/30/20 criterion, however, was determined to be a more realistic evaluation of the scores. If the subject did not meet

the criterion on the first test, a 1.0 indicating beginning reading level was assigned. For comparison, the readability levels of each cloze passage were utilized.

Ransom (1967) found significant correlations at the .01 level between the cloze test and informal reading inventory at the instructional and frustration reading levels for grade levels two to six. The correlations at the independent reading level were not significant at the .01 level of confidence for a majority of grades tested, the exception being grade four. In the first grade, correlations relevant to all three reading levels failed to achieve statistical significance.

The data were analyzed for Kirby's study by a simple analysis of variance, F ratios, t tests, and inspection. The .01 level of confidence was employed to determine significance of the F ratios and t tests between the means derived from the four instruments and the cloze test. The t test was utilized to determine significance of mean differences between any two sets of scores when the F test indicated rejection. Mean differences among the four instruments were significant for grades one through four, between sexes, low ability students, and able and less able readers. There were no significant differences among the mean scores on the instruments at the fifth and sixth grade levels, for high ability students or for outstanding readers. Mean scores on the cloze test did not differ from the Gilmore test at any level or for any subgroup. Mean scores on the cloze test did not differ significantly from those on the Gates Reading Tests except in grades one and two. Mean scores on the cloze test differed significantly from those on the Gray Oral Reading Test for the total sample, grade one, grade three, girls,

levels with a degree of accuracy comparable to that obtained with a properly constructed and administered informal reading inventory when the subjects are elementary school intermediate grade pupils. The cloze test scores correlated with the scores of the Gates-MacGinitie Reading Test and the Lorge-Thorndike Intelligence Test as well as the informal reading test.

In comparing the scoring criteria presented, a disparity is evidenced. At the independent reading level the interval is 50 to 62 percent or a range of 12 points. The largest interval is at the instructional level which is from 30 to 61 percent or a range of 31 points. An interval of 20 to 47 percent or a 27 point range for the frustration reading level is noted.

TABLE I

CLOZE SCORING CRITERIA - FIFTH WORD DELETION

INVESTIGATOR	CRITERIA		
	INDEPENDENT	INSTRUCTION	FRUSTRATION
Bormuth (1967)	50%	38%	
Bormuth (1968)	57%	44%	38%
Rankin & Culhane (1969)	61%	41%	
Ransom (1967)	50%	30%	20%
Alexander (1968)	62%	61-47%	47%

Scoring procedures have not been resolved as noted in research conducted by Gallant (1964) and Schoelles (1971). Gallant investigated two related problems: the validity and reliability of cloze tests as a measure of silent reading comprehension for children in grades one through three; and the effect of increased sentence length on the reading difficulty of passages designed for use with beginning readers. The first problem is relevant for this review.

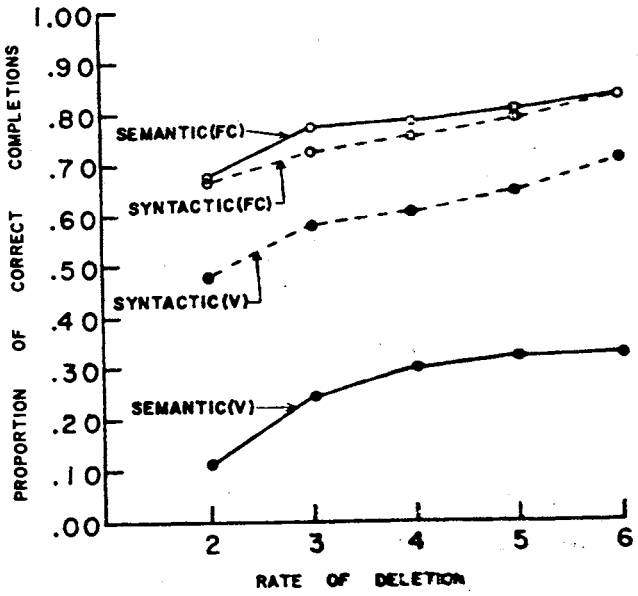
To determine validity of the cloze instruments as a measure of silent reading comprehension, a comparison was made between the ranking of pupils within each grade on a paragraph reading section of a standardized reading test, the Metropolitan Achievement Test, and the ranking on a cloze test. After a pilot study, a modified form of the cloze test was developed for administration to the first grades since the investigator found that these students could not handle the regular cloze procedure.

There were 273 pupils participating in the study conducted by Gallant (1964). After the administration of the standardized test, the cloze test which had been constructed from the Metropolitan Achievement Test was given. Two types of scoring were used; exact word and substitute word. Pearson product-moment correlation was used to determine if the rankings of pupils on the cloze test corresponded with the rankings on the standardized achievement test. Correlation coefficients for the exact scores for all three grades ranged from .65 to .81. The range of the correlations for the substitute scores of grades two and three was from .70 to .83. With the exception of the group comprised of the boys in grade three, the correlation coefficients for the substitute scores were slightly higher than the correlations for the exact

scores. For this group, Gallant found that the correlations for both sets of scores were the same. It was found that the correlations between the cloze tests and the standardized reading tests for each grade level in total and for each grade subdivided by sex were significant at the .01 level of confidence. An analysis of variance substantiated the validity of cloze scores as a measure of reading comprehension. The reliability coefficients for the cloze tests ranged from .90 to .97 for the exact scores and .95 to .97 for substitute scores, all significant at the .01 level of confidence. Use of substitute scores did not make any significant differences in the statistical relationships. It did decrease the efficiency and the objectivity of the scoring procedure as decisions to accept or reject specific responses became a matter of interpretation, dependent upon the judgment of the person doing the scoring.

The relationship between cloze test scores and individual reading ability were investigated by Schoelles (1971). Four hundred and seventy students in grades one through five were administered two reading subtests of the Stanford Achievement Test and cloze passages developed from several basal readers. At the third year level every fifth word was deleted while every seventh word was deleted at the other reading levels. Correlations on cloze passages at the grade mean ranged from .64 to .76 with the lower correlation being at the third level. A correlation of .72 was determined for all the subjects, grades one through six. Scoring was for the exact word, initially; however, synonyms were later included and analyzed. This was found to be particularly valuable with higher reading levels since words may become more precise and harder to predict.

Rate of deletions, or the manipulation of contextual constraints, has been considered in several studies. Fillenbaum, Jones, and Rapaport (1962) made a study which was concerned with the grammatical and lexical predictability of speech. The cloze procedure was used and every second, third, fourth, fifth, or sixth word was systematically deleted, depending on the experimental condition. The subjects were college students and were enrolled in introductory psychology classes. All subjects were tested in groups of 5 to 20. Fillenbaum defines form class predictability as the extent to which words are supplied of the same grammatical class as the missing item, i.e., the extent to which context allows prediction of the sort of word deleted. Verbatim predictability is the extent to which content would allow the exact word to be supplied. In comparing semantic (specifically informative items) with syntactic of function deletions, it was noted that performance improved between successive deletions. It appears that when the context allows prediction of the sort of word deleted (form class) the proportion of correct completions will be higher than when the subject must supply the precise prediction of the missing word (verbatim). In comparing semantic (specifically informative items) with syntactic of function deletions, it was noted that performance improved between successive deletions.



Source: S. Fillenbaum, L. V. Jones, & A. Rapoport, Journal of Verbal Learning and Verbal Behavior, 1963, 11, 186-194.

Figure 1. Proportion of Correct Verbatim (V) and Form-Class (FC) Completions: Semantic and Syntactic Items

Still another form of deletion was considered by McLeod (1965) with the standardization of GAP, a reading comprehension test. Fifteen passages, each of about 50 words length, were selected from various types of books for the cloze passages. In each passage the tenth word and then every eighth word were deleted. Two other mutilations of the same test were prepared, with various deletion patterns.

After the administration of the tests to first year college students, the final two forms of the GAP were developed. All of the children in grades two through eight in one system were administered the

two forms of the cloze test, the Watts' Reading Comprehension Test and the Schonell Silent Reading Test. Intercorrelations between the two forms of the cloze test, the Watts', and the Schonell Silent Reading Test ranged from .67 to .82. Retest correlations, after a three-month interval, ranged from .79 to .92. It was concluded that the GAP is a valid test of reading comprehension.

Potter (1968) answered several questions for the classroom teacher regarding the use of the cloze under usual classroom conditions for children of diverse reading abilities. He used two forms of deletions, every fifth word and every tenth word. Potter was also interested in determining if there was a relationship between comprehension scores derived from the experimental procedure and a standardized test of reading comprehension, the Standard Achievement Test. One hundred twenty-eight sixth-grade subjects were administered four cloze tests under two instructional conditions. In two samples, every fifth word was deleted and in two samples every tenth word was deleted. The passages were of fourth and eighth grade readability levels and were constructed so that two were "easy" and two were "difficult." Each subject took each test since on one set every word in five was deleted while on the other set every tenth word was deleted. The cloze tests yielded three scores: 1) a total percentage of the words correct, 2) percentage of content words correct, and 3) percentage of function words correct. Criterion for correctness used was the replacement of the exact word deleted. Potter theorized that a subject with a high function word score on a cloze test comprehends the structure of the passage while a high content word score may indicate familiarity with the vocabulary in question. On this basis, one would predict higher function word scores than

content word scores on passages of greater difficulty. Content words were defined as those words having meaning in themselves, like "mother," "tomorrow," and "car." Function words have little or no meaning other than the grammatical relationships they express, such as "the" and "of." From the resulting raw score, a percentage of correct function words and content words was calculated for each subject from each test. The first half of each test was scored separately from the second half, and a split-half reliability factor was then determined.

An analysis of variance design was used to examine the relationship between the two instruction treatments and the two reading levels. Passage difficulty and deletions were also examined. In order to examine differences between total cloze scores and those cloze measures derived from content word deletions and function word deletions, each type was scored independently and analyses reported separately. Variance associated with deletion rate was significant for total cloze scores, content word deletions, and function word deletions. In the case of the total scores and the function word deletions, means for the 1:10 deletion rate, were significantly higher ($<.01$). On content word scores, however, the 1:5 rate was higher ($<.01$). Since the 1:10 deletion rate provides a larger number of words, and therefore more context clues around each deletion, a larger number of correct responses was expected for this treatment. Differences in passage difficulty were found to be significant ($<.01$) for total scores, content word deletions, and function word deletions. "More able" and "less able" readers, provided the greatest single source of variance in cloze scores. Means were significantly different at the .01 level. A test-instructions by passage-difficulty interaction was found in content word scores at

the .01 level and total word scores at the .05 level. The test-instructions by deletion-rate interaction was significant ($<.05$) only for content words. Only the interaction between instructions for the 1:5 deletion rate showed significance for simple effects ($<.5$). This finding indicates that the correct completion of content word deletions requires somewhat different skills in the reader. The passage-difficulty by deletion rate interaction was significant at the .01 level for function words and for total cloze deletions. Tests for simple effects revealed that significant differences between means appear over the passage difficulty variables to favor the 1:10 by "easy" interaction. Interactions between the instructions, reading ability, and passage difficulty were significant for function words ($<.01$) and for content words (.05). A three-way interaction between instructions, reading ability, and deletion rate was significant ($<.01$) for the function word scores only. An examination of the correlations between the cloze tests and the Stanford Achievement Test scores revealed no difference in content word correlations and function word correlations for any of the between-subjects variables. Content word correlations, for the total sample, range from .72 to .83 while function word correlations range from .68 to .84. Each of these correlations, while significant in itself, shows no superiority of either content words or function words.

Summary

In summary the studies included in this chapter have indicated that the cloze procedure is a reliable measure of reading comprehension from the first grade through college level. The various studies have provided evidence of the reliability through positive correlations;

Taylor (1953) reported a coefficient of .88 while Gallant (1964) calculated coefficients ranging from .90 to .97 for exact scores and .70 to .83 for substitute scores. Bormuth (1962) computed reliabilities by two methods and obtained coefficients which ranged from .70 to .95. Three levels of reading were established through the cloze procedure and the informal reading inventory by Ransom (1967) and Alexander (1968). Several questions about the use of the cloze in the classroom were answered through Potter's (1968) study. McLeod (1965) developed and standardized a cloze test based on a tenth/eighth word deletion. Rankin (1958) concluded that the lexical form (nouns, verbs) was a more accurate measure of specific reading comprehension than general reading skill; however, Alexander (1968) felt that both lexical and structural language factors are utilized with the cloze test and are both important components of reading comprehension.

CHAPTER III

DESIGN AND METHODOLOGY

Introduction

The purpose of this study was to validate a reading instrument, the Hicks Cloze-Reading Test. The instruments used were the Cloze Tests, constructed specifically for this study, a standardized informal reading inventory, and a standardized reading test.

The study was designed to compare scores of the Hicks Cloze-Reading Test with the independent, instruction, and frustration reading levels of a standardized informal reading inventory to determine whether or not there were significant relationships. To facilitate the analysis of this problem, three major questions were developed for investigation.

1. Is there a relationship between the grade equivalency of the Hicks Cloze-Reading Test and the Standard Reading Inventory independent reading level at the second, third, fourth, fifth, and sixth grade levels?

2. Is there a relationship between the grade equivalency of the Hicks Cloze-Reading Test and the Standard Reading Inventory instructional reading level at the second, third, fourth, fifth, and sixth grade levels?

3. Is there a relationship between the grade equivalency of the Hicks Cloze-Reading Test and the Standard Reading Inventory frustration reading level at the second, third, fourth, fifth, and sixth grade levels?

The Population and Data Collection Procedure

There were 494 students tested in two school systems and of this number 347 subjects were used for this study. In Chandler, a central Oklahoma community, 334 students were screened, while in Perkins, a north central Oklahoma community, 160 students were screened. Although these school populations were predominantly white, there were a small percentage of black and Indian children. The subjects were from various socio-economic levels. Complete data were obtained for 347 subjects. The high attribution was the result of the need to determine three reading levels, independent, instructional, and frustration, for each subject on one of the test instruments, the Standard Reading Inventory. Due to absentism and moving, data was incomplete for 17 of the subjects. All of the testing was completed in one school system before it was begun in the other school system. Approximately four months were spent in testing all of the subjects.

TABLE II

SCREENING POPULATION REPRESENTATION

N = 494	Chandler	Perkins	Number of Subjects
Second grade	60	41	101
Third grade	56	33	89
Fourth grade	60	46	106
Fifth grade	69	29	98
Sixth grade	89	11	100

The Gates-MacGinitie Reading Test was administered to all the subjects in Chandler by the investigator and the reading teacher and in Perkins by the investigator. Administration of the test was according to directions found in the test manual. A vocabulary and silent reading comprehension score were obtained and recorded for each subject. These scores later gave direction for beginning the Standard Reading Inventory, an individualized informal reading inventory. All scoring followed instructions in the test manual.

Prior to administering the Standard Reading Inventory (SRI), a training session was held for those who were to do the testing. A demonstration, allowing for questions, was made by the investigator. Each of the four examiners who were assisting in Chandler then administered the SRI to several children. The additional examiners who assisted in Perkins were all qualified reading specialists from Oklahoma State University, already familiar with the test.

The Standard Reading Inventory was then administered on an individual basis to establish the independent, instructional, and frustration reading levels of each subject. Scoring procedures followed instructions in the test manual. After each test was rechecked and evaluated for accuracy by the investigator, the derived grade levels were recorded on the cover of the test booklet. Since, in this study, an exact reading level was necessary, readabilities were made on the eight passages of the SRI at the intermediate levels, utilizing the Dale-Chall Readability formula. The oral and silent readability levels established were then averaged to determine one reading level for each of the fourth, fifth, sixth, and seventh grade passages.

The Hicks Cloze-Reading Test was then administered to all of the subjects, regardless of whether three reading levels had been obtained on the Standard Reading Inventory. Cloze tests were compiled for each child on an individual basis utilizing the instructional reading level established on the Standard Reading Inventory. Since two forms of the Hicks Cloze-Reading Test were to be administered, the same procedure was followed for both in the classroom. One-half of the class began with Form I and the other half began with Form II. Direction sheets were the same for both forms. After each subject had completed the first form, the other form was then given to him to complete. The test was not timed. Any tests that were not completed with the examiners present were completed under the supervision of the classroom teacher. This prevented the slower reader from being penalized by a time element.

After the Gates-MacGinitie Reading Test, the Standard Reading Inventory, and Hicks Cloze-Reading Test had been administered, complete information was available for 347 subjects. This is divided by school and grade level below.

TABLE III

NORMING POPULATION REPRESENTATION

N = 347	Chandler	Perkins	Number of Subjects
Second grade	25	23	48
Third grade	39	28	67
Fourth grade	44	39	83
Fifth grade	52	25	77
Sixth grade	61	11	72

Rationale for Cloze Instrument

In developing a measuring instrument, specifically a cloze test, several criteria must be considered, i.e., type of deletions, scoring, deletion rate, level of administration. Some of the research most pertinent to this concept will be reviewed.

Taylor (1957) studied correlations between three types of cloze tests formed by deleting 1) any-words, 2) easy-words (i.e., conjunctions, pronouns, articles, verb auxiliaries), and 3) hard words (i.e., nouns, verbs, adverbs). Significant correlations at the .001 level were found between all three types of cloze tests and various criterion tests. Easy-word correlations were the smallest and with one exception, the any-word correlations were the highest. Taylor concluded that for testing comprehension, aptitude, and readability, the any-word form of deletion was superior.

Studies were made by Taylor (1953), Rankin (1958), and Ruddell (1963), utilizing two types of scoring. Each investigator made cloze tests over a set of passages and obtained a set of scores by counting the responses exactly matching the deleted words and another set including the responses that were synonymous with the deleted words. The investigators found that including synonyms in the scores increased the variances of the tests. Ruddell also found that including synonyms slightly increased the correlations with scores on a reading achievement test. Hafner (1964) determined a correlation of .61 between scores found by counting exact word deletions and scores obtained by counting responses that did not match the deleted word but which were grammatically correct. Bormuth (1965) devised a study whereby cloze test results

were classified according to their semantic and grammatical relationships to the deleted word. Scores based on each of the seven answer classification groups were examined to determine the most reliable method of scoring the cloze test. The seven scores were then correlated with the total reading score on the achievement test. Scores obtained by counting the EGC (exact word, grammatically correct) responses were superior to any of the other types of scores included in this study. Bormuth concluded that when cloze tests are used as measures of individual differences in reading ability, scores obtained by counting responses exactly matching the deleted words seem to yield the most valid scores. A second conclusion was that when cloze tests are used in readability studies as measures of the comprehension difficulties of passages, scores obtained by counting responses exactly matching the deleted words seem to yield the greatest amount of discrimination among passage difficulties. From the above findings, it would appear that responses exactly matching the deleted words furnish the most valid measures of comprehension.

Research investigating the use of cloze tests as a measure of comprehension ability was conducted by Bormuth (1962), Gallant (1964), and Potter (1968). Bormuth investigated three aspects of the cloze test: 1) their validities as measures of comprehension ability, 2) their efficiencies when used for the purposes of discriminating among the difficulties of tests and the abilities of individuals, and 3) their validities as measures of the amounts of comprehension with which passages are read. Nine passages, divided by subject matter and level of difficulty, were administered to subjects in the intermediate grades. The results supported the hypothesis that cloze tests are valid measures

of comprehension ability. Scores on the cloze tests correlated with scores on the comprehension test scores administered to the subjects. The hypothesis was supported that cloze tests are efficient when used for the purpose of measuring readability and comprehension skills. Gallant (1964), in one part of a study, investigated the validity and reliability of cloze tests as a measure of silent reading comprehension for pupils in grades one through three. An analysis of variance substantiated the validity of the cloze scores as a measure of reading comprehension. Similar results were obtained by Potter (1968) when total cloze scores correlated significantly with standardized reading test scores. The cloze test procedures were identified which yielded a valid and reliable measure of reading comprehension for upper elementary children.

Much research utilizing cloze procedure with upper elementary, junior and senior high levels, and college-age subjects has been conducted. Gallant (1964), Kirby (1967), Ransom (1967), and Schoelles (1971) have reported successful research utilizing cloze with lower elementary level children. Ransom also established a cloze criteria based on a study conducted with grades one through three. This would tend to indicate that cloze can be used successfully with grades one through college level to measure reading comprehension.

The research contains studies that have established scoring criteria on every fifth word deletion (Bormuth, 1968; Ransom, 1965; Alexander, 1968). McLeod (1965) standardized a cloze test, GAP, which was based on a ten/eight word deletion system and was normed on aggregate scores. Sauer (1969) used a tenth word deletion; however,

scoring was made from criteria established on a five word deletion system.

A number of studies using the tenth word deletion system have been completed. Taylor (1953) measured the effect of deletions per passage in ranking the readability of selected passages. A later study by Taylor (1957) tested the hypothesis that cloze scores would correlate significantly on specifically constructed tests as well as intelligence and aptitude tests. Schneyer (1965) explored the effects of the cloze procedure upon the reading comprehension of sixth grade students. The tenth word deletion was much more highly related to intelligence than was the noun-verb deletion system (.63 vs. .42). This substantiates Rankin's (1958) contention that the every nth deletion system is more related to intelligence, whereas the selective deletion of nouns and verbs provides a measure of comprehension less influenced by intelligence. Potter (1968) using two deletions systems; every fifth word and every tenth word, determined the relationship of content and function words to the surrounding context.

Studies by Aborn (1959) and MacGinitie (1961) suggest that an increase in cloze scores occurs as the deletion rate increases. It is also reported in these studies that the amount of increase is curvilinear and that the rate of acceleration is negative as the number of words of surrounding context is increased. MacGinitie's investigation, also, indicated that context extending beyond 16 to 24 words between deletions does not contribute to an increase in cloze scores. It has been shown that the semantical and grammatical factors combine to strengthen the redundancy of a passage through the contextual constraints. Context containing between 5 to 10 words between deletions

appears to be the most effective toward the accuracy of word prediction (Aborn, 1959). There is a need for a scoring criteria based on an every tenth word deletion. This need has been expressed by many of the investigators of the cloze procedure (Taylor, Alexander, Gallant, and Ransom).

Construction of the Hicks Cloze-Reading Test

Utilization of this research as well as other studies reviewed, led to the formation of the Hicks Cloze-Reading Test used in this study. The instrument contains fifteen stories, ranging in readability from 2.0 to 9.5. Readability levels were established through the use of the Spache and Dale-Chall Readability formulas. These readability formulas have been established as reliable measures of the readability of materials. Scoring criterion was determined to be the exact word replacement for one point, making a total cloze score for each passage. The deletion pattern was structural (any-word) and was an every tenth word deletion system. There were no deletions in the first or the last lines of any passage. Each passage was approximately 200 words in length. Passages with a higher readability were somewhat longer. Stories of interest to young children were designed by the investigator. Each primary level passage from 2.0 to 3.5 was prepared utilizing the Spache Readability formula (1953). At these levels, the stories were written so as to have a continuous theme of adventure, one story idea being about twins (Form I) and the story idea being about two boys (Form II). Passages at the intermediate levels, 4.0 to 9.5, were measured by the Dale-Chall Readability formula (1948). These selections were chosen from various publications.

Reliability of Hicks Cloze-Reading Test

Reliability of the Hicks Cloze-Reading Test was determined through the administration of the instrument to 494 subjects in two central Oklahoma communities. Alternate forms of the test were given each group. Half of the class began with Form I; the other half of the class began with Form II. The procedure was then reversed to accomodate for any transfer effects. Both forms were administered on the same day to accomodate the time interval. Reliability was strengthened through the construction and administration of the instrument. Pearson-Product Moment correlations were made using the cloze percentages at the independent, instructional, and frustration levels for each grade to determine the reliability of the alternate forms. This information is presented in Table IV. All correlations were significant; however, the

TABLE IV

ALTERNATE FORM CORRELATIONS FOR HICKS CLOZE-READING TEST, FORMS I AND II USING CLOZE PERCENTAGE CRITERIA

Grade	Independent Level	Instructional Level	Frustration Level
Second	.35**	.40***	.46*
Third	.38*	.75*	.79*
Fourth	.66*	.60*	.60*
Fifth	.59*	.69*	.70*
Sixth	.60*	.75*	.79*

*Significant at .01 level of confidence
**Significant at .05 level of confidence
***Significant at .02 level of confidence

second grade independent correlation was significant at the .05 level of confidence, while the instructional correlation of .40 was significant at the .02 level of confidence. The remaining correlations were all significant at the .01 level of confidence. This would support the reliability of the two forms of the Hicks Cloze-Reading Test.

Validity of Hicks Cloze-Reading Test

Content validity was built into the test through the application of established readability formulas to each story. The test, as it is constructed, is built upon a rationale from the research. Concurrent validity was determined through significant Pearson-Product Moment correlations established between the Standard Reading Inventory and each form of the Hicks Cloze-Reading Test. Correlations at the .01 level were obtained at each reading level; independent, instructional, and frustration for grades two to six for Form I. Significant correlations at the .01 level were obtained for Form II at each of the reading levels, with one exception. The third grade independent level correlation was not significant. This would support the validity of the two forms of the Hicks Cloze-Reading Test.

Instruments Used in the Study

Gates-MacGinitie Reading Tests, Primary B, Primary C, and Survey D: these reading tests are designed to be used with children in grades two to six. An examination of the technical manual indicates that the test has high reliability, but no data on validity has been reported. Alternate form reliability coefficients on the comprehension sub-test at the intermediate level was reported as ranging from .83 to .89.

Split half reliability coefficients on the same sub-test and levels were reported as ranging from .94 to .96. The consistency of measurement is evident on the basis of these reports.

The authors of the Gates-MacGinitie Reading Tests have made provisions for converting raw scores into scores corresponding to those of the 1958 edition of the Gates Reading Tests. The Gates-MacGinitie Reading Tests are designed to measure three reading components: 1) speed and accuracy, 2) vocabulary, and 3) comprehension. The speed and accuracy scores, however, were omitted in this investigation.

Standard Reading Inventory, developed by Robert McCracken (1964): the Standard Reading Inventory was designed to be administered to individual subjects, measuring the independent, instructional, and frustration levels. The Spache and the Dale-Chall Readability Formulas were used to determine the reading levels which in turn were based on basal reading series.

Form B of the Standard Reading Inventory consists of 11 stories for oral reading, eight stories for silent reading, and 11 word lists for measuring skill in pronouncing words in isolation.

The reading achievement areas for measurement in the Standard Reading Inventory are recognition vocabulary with words both in isolation and in context, oral reading errors, comprehension, and speed. Comprehension included recall after oral and silent reading and interpretation and word meaning after oral and silent reading.

Concurrent validity of the Standard Reading Inventory and the California Reading Test was ascertained with 79 children completing second grade. The correlation was .87. A second study was made with the two sub-tests of the Stanford Achievement Test and the correlations

were .77 with the Standard Reading Inventory and Stanford Comprehension and .88 with the Standard Reading Inventory and Stanford Word Meaning.

Scoring

Raw scores on the Gates-MacGinitie Reading Test were converted to grade equivalents according to the test manual. In order to obtain each of the three reading levels from the Standard Reading Inventory, it was necessary to make readabilities on the oral and silent passages for each of the intermediate levels. These readabilities (according to the Dale-Chall Readability formula) were then averaged to obtain a reading level. These were determined to be: 1) Fourth level - 4.8; 2) Fifth level - 5.6; 3) Sixth level - 6.7; and 4) Seventh level - 7.7. These grade equivalents were then recorded for each subject. Cloze passages were scored for exact word replacement. Misspellings were allowed unless similarity to the original word was questionable. The number of words exactly replaced made up the cloze score for that passage.

Since cloze scores were recorded as the number correct on each passage, it was essential, prior to analysis, to convert them to percentages in order to equate them to the criteria of the informal reading inventory. Since no criteria are available for a tenth word deletion, the present investigation has developed some.

To arrive at the criteria identifying the independent, instructional, and frustration reading levels, the following procedure was employed. All of the 347 subjects' scores were grouped according to the comprehension score obtained on the Gates-MacGinitie Reading Test. The Hicks Cloze-Reading Test raw score for the corresponding passage on each form was divided by the number of deletions on the passage and was

cumulative for each grade level. The total was then divided by the number of scores on that level to determine the mean percentage. The mean instructional percentage obtained from Form I of the Hicks Cloze-Reading Test was 45.7 while the mean instructional percentage for Form II was 43.8. An interval of five points on either side of the mean was utilized as a basis in establishing a range. Each of the instructional

TABLE V

PERCENTAGE CRITERIA FOR IDENTIFICATION OF INDEPENDENT, INSTRUCTIONAL AND FRUSTRATION READING LEVELS ON HICKS CLOZE-READING TESTS, FORMS I AND II

	% Cloze I	Range	% Cloze II	Range
Independent Level		52 above		50 above
Instructional Level	45.7	41 - 51	43.8	39 - 49
Frustration Level		40 below		38 below

mean percentages are midpoint in the range. Therefore, a cloze percentage score falling between 41 to 51 on the Form I and between 39 to 49 on Form II would establish an instructional level. A cloze percentage above 52 on the Form I and above 50 on the Form II would establish independent levels for those forms while a cloze score below 40 on Form I and 38 on Form II would identify the frustration reading level.

Statistical Treatment of Data

Pearson Product-Moment correlations were made between the scores obtained on the Standard Reading Inventory and the Hicks Cloze-Reading Test at the independent, instructional, and frustration reading levels for grades 1, 2, 3, 4, 5, and 6. Correlations were made between the Gates-MacGinitie Reading Test comprehension scores and the instructional level obtained on the Standard Reading Inventory. Correlations were made between the two forms of the Hicks Cloze-Reading Test to determine reliability between the two forms. The .01 level of confidence was utilized to determine level of significance in all correlations.

CHAPTER IV

ANALYSIS OF DATA

The purpose of this study was to validate an evaluative reading instrument, the Hicks Cloze-Reading Test. This purpose could only be established if an instrument built on the cloze procedure can be expected to provide an examiner with certain information, more specifically the independent, instructional, and frustration reading levels of the individual. At present, this information is secured from individually administered informal reading inventories.

The nature of the data delineated the statistical methods employed in the analyses. Pearson's formula for obtaining a correlation coefficient as a measure of the degree of relationship between two variables was selected. This analysis assumes that the two variables are linearly related. To determine the significance of r, the appropriate table (Bruning and Kintz) was entered at the .01 level of confidence.

To facilitate in the analysis of the problem three major questions were developed for the study.

1. Is there a relationship between the grade equivalency of the Hicks Cloze-Reading Test and the Standard Reading Inventory independent reading level at the second, third, fourth, fifth, and sixth grade levels?

Correlations were made between the Standard Reading Inventory and Form I of the Hicks Cloze-Reading Test at each grade level; then were made between the Standard Reading Inventory and Form II of the Hicks

TABLE VI

RELATIONSHIP OF HICKS CLOZE-READING TEST, FORMS I AND II, AND STANDARD READING INVENTORY (SRI) AT INDEPENDENT READING LEVEL GRADES TWO-SIX

Instruments	N	Mean	SE Mean	SD	Pearson r	Level of Significance
Second Grade:						
SRI	39	2.150	.088	.553	.563	.01
Cloze I	39	2.887	.084	.530		
SRI	34	2.210	.091	.532	.535	.01
Cloze II	34	2.494	.092	.539		
Third grade:						
SRI	54	2.712	.115	.846	.394	.01
Cloze I	54	3.392	.094	.695		
SRI	53	2.836	.100	.728	.162	NS
Cloze II	53	2.944	.112	.822		
Fourth grade:						
SRI	75	3.085	.114	.990	.548	.01
Cloze I	75	3.497	.094	.820		
SRI	74	3.129	.110	.947	.497	.01
Cloze II	74	3.282	.111	.963		
Fifth grade:						
SRI	68	3.036	.109	.907	.394	.01
Cloze I	68	3.479	.122	1.008		
SRI	68	3.049	.116	.906	.511	.01
Cloze II	68	3.511	.127	1.049		
Sixth grade:						
SRI	71	3.411	.093	.791	.289	.01
Cloze I	71	3.754	.075	.637		
SRI	70	3.407	.094	.792	.287	.01
Cloze II	70	3.754	.114	.959		

Cloze-Reading Test to determine an answer for this question. This data is presented in Table VI. At the fifth grade level the number of subjects remained the same for each correlation; however, at each of the other grade levels there were unequal N's. Some of the subjects established an instructional and/or frustration level on the Hicks Cloze-Reading Test but did not establish an independent level.

Correlations, ranging from .289 to .563 were established between the Standard Reading Inventory and the Hicks Cloze-Reading Test, Form I, and were significant for the independent reading level at the .01 level of confidence for grades two, three, four, five, and six. Significant correlations were noted in Table VI, ranging from .287 to .535, were established for Form II of the Hicks Cloze-Reading Test at each grade level with one exception, grade three, which had a non-significant correlation of .162. Correlations obtained for Form I were slightly higher than for Form II at all grade levels except five where the correlation for Form II was higher (.511 vs. .394).

The standard deviations for the cloze tests were the smallest at grade two with a gradual increase becoming evident at each successive grade level. A full deviation unit was reached at the fifth grade level for both forms; however, the sixth grade deviations were again lower. Similar results were noted in Table VI in examining the standard error of the means. At both the second and sixth grade levels the standard errors were small while at the fifth grade level the range was from .109 to .127, indicating a wider interval.

From the results obtained, a significant relationship between the SRI and the Hicks Cloze-Reading Test at the independent level for grades two-six was established. The highest correlations were established at

the second grade level, and the lowest correlations were established at the sixth grade level for both forms.

The second question investigated in the study was:

2. Is there a relationship between the grade equivalency of the Hicks Cloze-Reading Test and the Standard Reading Inventory instructional reading level at the second, third, fourth, fifth, and sixth grade levels?

Correlations ranging from .387 to .599 were obtained between the Standard Reading Inventory instructional level and the Hicks Cloze-Reading Test, Form I, for grades two-six. This data is presented in Table VII. Correlations ranging from .393 to .631 were obtained between the SRI instructional level and the Hicks Cloze-Reading Test, Form II, for grades two-six. All of the correlations were significant at the .01 level of confidence. The highest correlations for both forms were established at the fifth grade level (.599-.631) while the lowest correlations (.430-.393) were established at the sixth year level.

Standard deviations extended from .616 at the second grade level to 1.441 at the sixth grade level. All of the standard deviations were between one-half and one and one-half deviation units. At the fifth and sixth grade levels this would include from 68 to 82 percent of the students tested at those levels.

At the instructional reading level, as noted in Table VII, the range in the standard error of the mean was small. In comparing the two forms of the Hicks Cloze-Reading Test, smaller errors of the mean and standard deviations were established for Form I. It was noted that at the second grade level, the means for both instruments were at midpoint of the third year and that at the third year level, the means for both of the instruments utilized in the study were at the fourth grade

TABLE VII

RELATIONSHIP OF HICKS CLOZE-READING TEST, FORMS I AND II, AND STANDARD READING INVENTORY (SRI) AT INSTRUCTIONAL READING LEVEL GRADES TWO-SIX

Instruments	N	Mean	SE Mean	SD	Pearson r	Level of Significance
Second grade:						
SRI	45	3.405	.120	.805	.387	.01
Cloze I	45	3.604	.091	.616		
SRI	36	3.555	.137	.823	.483	.01
Cloze II	36	3.305	.128	.772		
Third grade:						
SRI	63	4.096	.142	1.133	.557	.01
Cloze I	63	4.147	.110	.878		
SRI	58	4.229	.155	1.181	.513	.01
Cloze II	58	3.970	.126	.963		
Fourth grade:						
SRI	80	4.639	.144	1.296	.560	.01
Cloze I	80	4.440	.108	.969		
SRI	75	4.726	.145	1.258	.582	.01
Cloze II	75	4.294	.123	1.068		
Fifth grade:						
SRI	74	4.697	.141	1.213	.599	.01
Cloze I	74	4.533	.144	1.240		
SRI	66	4.828	.143	1.169	.631	.01
Cloze II	66	4.687	.181	1.471		
Sixth grade:						
SRI	72	5.313	.126	1.077	.430	.01
Cloze I	72	5.061	.167	1.421		
SRI	70	5.360	.119	1.003	.393	.01
Cloze II	70	5.292	.172	1.441		

level. The fourth and fifth grade means were at mid-point of the fourth year and the sixth grade means were at the low fifth grade level. Utilizing the standard error of the means would not increase the dispersion of the scores; however, the standard deviations would raise and/or lower them considerably. Possible implications of this data will be discussed later in this study.

The third question developed for analysis in this study is concerned with the frustration reading level.

3. Is there a relationship between the grade equivalency of the Hicks Cloze-Reading Test and the Standard Reading Inventory frustration reading level at the second, third, fourth, fifth, and sixth grade levels?

Significant correlations, ranging from .422 to .635 were established at the .01 level of confidence for grades two-six at the frustration reading level. This data is presented in Table VIII. Correlations of .422 to .635 were established between the Hicks Cloze-Reading Test, Form I, and the Standard Reading Inventory. Somewhat higher correlations (.458 to .634) were found between the Hicks Cloze-Reading Test, Form II, and the frustration reading level of the Standard Reading Inventory. The highest correlation between the two instruments were found at the third grade level while the lowest correlations were at the sixth grade level.

Mean scores for the fifth grade subjects on the Standard Reading Inventory indicated frustration at the mid-point of the fifth year level of the fourth year on both forms of the Hicks Cloze-Reading Test. Sixth grade students frustrated at the lower limits of the sixth grade on the Standard Reading Inventory and at the mid-point of the fifth year level on both forms of the Hicks Cloze-Reading Test. The standard error of

TABLE VIII

RELATIONSHIP OF HICKS CLOZE-READING TEST, FORMS I AND II, AND STANDARD READING INVENTORY (SRI) AT FRUSTRATION READING LEVEL GRADES TWO-SIX

Instruments	N	Mean	SE Mean	SD	Pearson r	Level of Significance
Second grade:						
SRI	48	4.071	.156	1.087	.539	.01
Cloze I	48	4.033	.119	.825		
SRI	48	4.007	.157	1.090	.607	.01
Cloze II	48	3.352	.140	.970		
Third grade:						
SRI	66	4.811	.176	1.431	.635	.01
Cloze I	66	4.553	.131	1.068		
SRI	66	4.803	.177	1.442	.634	.01
Cloze II	66	4.172	.140	1.144		
Fourth grade:						
SRI	81	5.494	.155	1.399	.585	.01
Cloze I	81	4.938	.116	1.050		
SRI	82	5.412	.145	1.314	.625	.01
Cloze II	82	4.418	.119	1.081		
Fifth grade:						
SRI	77	5.525	.152	1.336	.607	.01
Cloze I	77	4.994	.153	1.343		
SRI	76	5.496	.151	1.323	.623	.01
Cloze II	76	4.678	.178	1.558		
Sixth grade:						
SRI	70	6.226	.139	1.167	.422	.01
Cloze I	70	5.544	.148	1.242		
SRI	72	6.222	.137	1.170	.458	.01
Cloze II	72	5.613	.177	1.510		

the mean at the fifth year was somewhat larger than at the sixth year level.

Standard deviations were predominately grouped above the one deviation unit, with the exception of the two Hicks Cloze-Reading Tests at the second year grade level. These deviations, noted in Table IX, were .825 and .970 respectively for Forms I and II.

To further substantiate the cloze percentage scores reported in Table V, the same computational procedures were followed using the criterion instrument, the Standard Reading Inventory and both forms of the Hicks Cloze-Reading Test. The SRI provided three reading level scores, independent, instructional, and frustration. Utilization of these scores and the raw cloze scores provided a comparable cloze scoring criteria and range as presented in Table IX. This data supports

TABLE IX

PERCENTAGE CRITERIA FOR IDENTIFICATION OF INDEPENDENT, INSTRUCTIONAL AND FRUSTRATION READING LEVELS ON HICKS CLOZE-READING TESTS, FORMS I AND II, DETERMINED BY STANDARD READING INVENTORY

	% Cloze I	Range	% Cloze II	Range
Independent Level	60.2	54-100	54.7	50-100
Instructional Level	44.6	41-53	42.7	40-49
Frustration level	35.4	0-40	35.2	0-39

the cloze percentage criteria established using comprehension scores from the Gates-MacGinitie Reading Tests.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This study was concerned with the validation of the Hicks Cloze-Reading Test, an evaluative instrument to be used by the classroom teacher in assessing the reading ability levels of children. Three instruments, the Gates-MacGinitie Reading Test, the Standard Reading Inventory, and the Hicks Cloze-Reading Test, Form I and II, were administered to 494 students, in grades 2-6, in two central Oklahoma communities. Due to insufficient data, 347 subjects were used for this study.

The Gates-MacGinitie Reading Test was utilized as a screening instrument and as a basis for the establishment of the cloze percentage scores. The scores obtained on this instrument gave direction for the individually administered Standard Reading Inventory (SRI), two forms of the Hicks Cloze-Reading Test were then compiled and administered to each class of students. The latter test was not timed.

Scoring criteria were established across reading levels for the Hicks Cloze-Reading Test since there is a need for a scoring criteria built on an every tenth word deletion system. The mean cloze percentages established for the instructional reading level were 45.705 for Form I and 43.809 for Form II. An interval of five points in each direction from the mean established an instructional range of 41 to 51 for Form I. The independent level for Form I would be any score above 52

while the frustration reading level would be any score below 40. The instructional range established for Form II ranged from 39 to 49. The independent reading level is determined by a score above 50 while the frustration level is any score below 38.

To determine if a relationship existed between the SRI and each form of the Hicks Cloze-Reading Test, Pearson Product-Moment correlations were made at each grade level 2-6 for the independent, instructional, and frustration reading levels. At the independent level, correlations for Form I, ranging from .289 to .563, were significant at the .01 level of confidence. Significant correlations, ranging from .287 to .535 were obtained for Form II at each grade level except grade three. At the instructional reading level, correlations ranging from .387 to .599 were obtained for Form I while correlations ranging from .393 to .631 were obtained for Form II. All of the correlations were significant at the .01 level of confidence. Correlations, ranging from .422 to .635, were established for Form I at the frustration reading level. The correlations for Form II ranged from .458 to .634 and were all significant at the .01 level of confidence.

Correlation coefficients between the alternate forms of the Hicks Cloze-Reading Test, Forms I and II, ranged from .35 at the second grade independent level to .79 at the sixth grade frustration level. Significant correlations at the .01 level or greater were established for each of the reading levels; independent, instructional, and frustration for grades three, four, five, and six. Second grade independent and instructional reading level correlations were significant but not at the .01 level.

An examination of the data presented in Tables VI, VII, and VIII indicated that the correlations at the sixth grade level, although significant, were lower than the other four grade levels. This could indicate a weakness of one or both of the instruments at this level. Utilization of the scoring criteria of the Standard Reading Inventory could have prevented sixth year students from achieving an accurate reading placement since a time criteria, at this level, may have been a delimiting factor.

This data would tend to indicate that the Hicks Cloze-Reading Test, built upon an every tenth word deletion, can correctly identify the independent, instructional, and frustration reading levels of children in grades 2-6.

Conclusions

1. A new scoring criteria built on a tenth word deletion can be successfully established.

2. The Hicks Cloze-Reading Test, Forms I and II, are a valid and reliable measure of the independent reading level for pupils in grades 2 to 6. Reliability coefficients, although low, ranged from .35 to .60 and were all significant. Correlations at the independent level between the SRI and the Hicks Cloze-Reading Tests ranged from .289 to .563 for Form I and .162 to .535 for Form II.

3. The Hicks Cloze-Reading Test, Forms I and II, are valid and reliable measures of the instructional reading level for pupils in grades 2-6. Significant reliability coefficients, ranging from .40 at the second grade level to .75 at the sixth grade level, were established.

Correlations at the instructional level between the SRI and the Hicks Cloze-Reading Tests ranged from .387 to .599 for Form I and from .483 to .631 for Form II.

4. The Hicks Cloze-Reading Test, Forms I and II, are valid and reliable measures of the frustration reading levels for pupils in grades 2-6. Significant reliability coefficients were established at the frustration reading level ranging from .46 to .79. Correlations, significant at the .01 level, between the SRI and the Hicks Cloze-Reading Test were established ranging from .539 to .635 for Form I and from .458 to .634 for Form II.

5. The Hicks Cloze-Reading Test, Form I or Form II, will, for purposes of evaluation, produce valid and reliable measures of pupil reading ability.

Recommendations

1. The Hicks Cloze-Reading Test is recommended as a valid and reliable measure of reading ability in the upper primary and intermediate grades. This recommendation is based upon the significant correlations established between the reading levels of the Standard Reading Inventory and each form of the Hicks Cloze-Reading Test for grades 2-6.

2. This investigator recommends that the classroom teacher be cautious in establishing a reading range for students, since if too wide an interval exists, the child may inadvertently be placed in a frustrating learning situation. Scores on the Hicks Cloze-Reading Test or on any other evaluative instrument should be interpreted and evaluated as only a part of the total learning picture for an individual. The reading abilities of each child must be considered in establishing the

reading program. Problem areas should be properly evaluated and corrected in order to provide a satisfactory learning environment.

3. This investigator recommends that additional testing, utilizing the Hicks Cloze-Reading Test, be conducted with school populations containing a more equal distribution of racial differences. Before this instrument or any other evaluative instrument is administered, the investigator must determine what, if any, dialectial barriers exist and to what extent these barriers will effect the scores obtained.

A SELECTED BIBLIOGRAPHY

Aborn, M., Rubenstein, H., & Sterling, T. D. Sources of contextual constraint upon words in sentences. Journal of Experimental Psychology, 1959, 57, 171.

Alexander, H. W. An investigation of the cloze procedure as a measuring device designed to identify the independent, instruction, and frustration reading levels of pupils in the intermediate grades. Unpublished doctoral dissertation, University of Illinois, 1968.

Betts, E. A. Foundations in Reading Instruction. New York: American Book Company, 1957.

Bormuth, J. R. Cloze tests as measures of readability and comprehension ability. Unpublished doctoral dissertation, University of Indiana, 1962.

Bormuth, J. R. Validities of grammatical and semantic classifications of cloze test scores. In Figurel, J. A. (Ed.), Reading and Inquiry. Proceedings of the International Reading Association, 1965, 10.

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

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

Botel, M. Newer techniques in the supervision of reading instruction. Frontiers of Elementary Education. Syracuse, New York: Syracuse University Press, 1957.

Bruning, J. L. & Kintz, B. L. Computational Handbook of Statistics. Dallas: Scott-Foresman Book Company, 1968.

Dale, E. & Chall, J. S. A formula for predicting readability. Educational Research Bulletin, 1948, 27, 11-20.

Fillenbaum, S., Jones, L. V., & Rapoport, A. The predictability of words and their grammatical classes as a function of rate of deletion from a speech transcript. Journal of Verbal Learning and Verbal Behavior, 1963, 11, 186-194.

Gallant, R. An investigation of the use of cloze tests as a measure of readability of materials for the primary grades. Unpublished doctoral dissertation, University of Indiana, 1964.

Gallent, R. Use of Cloze tests as a measure of readability in the primary grades. In Figurel, J. A. (Ed.), Reading and Inquiry. Proceedings of the International Reading Association, 1965, 10.

Goodlad, J. I. & Anderson, R. H. The Non-Graded Elementary School. New York: Harcourt, Brace & World, 1963, 27-28.

Hafner, L. E. Relationships of various measures to the cloze. In Thurston, E. L. & Hafner, L. E. (Eds.), Thirteenth Yearbook of the National Reading Conference. Milwaukee: The National Reading Conference, Inc., 1964, 135-145.

Jongsma, E. The Cloze Procedure as a Teaching Technique. International Reading Association, 1971.

Killgallon, P. A. A study of relationships among certain pupil adjustments in reading situations. Unpublished doctoral dissertation, Pennsylvania State University, 1942.

Kirby, C. L. A comparison of scores obtained on standardized oral and silent reading tests and a cloze test. Unpublished doctoral dissertation, Ball State University, 1967.

Kirby, C. L. Using the cloze procedure as a testing technique. April, 1968, Ed- 019-ZOZ.

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

McCracken, R. A. Standardized reading tests and informal reading inventories. Education, February, 1962, 82.

McLeod, J. GAP Reading Comprehension Test. Melbourne, Australia: Heinemann, 1965.

McLeod, J. & Anderson, J. Readability assessment and word redundancy of printed English. Psychological Reports, 1966, 18, 35-38.

Miller, G. A., Galanter, E., & Pribram, K. H. Language and Communication. New York: McGraw-Hill Book Company, 1951.

Miller, G. A., Galanter, E., & Pribram, K. H. Plans and the Structure of Behavior. New York: Henry Holt and Company, 1960.

Millsap, L. N. A study of teachers' awareness of frustration reading level among their pupils in basal readers. Unpublished doctoral dissertation, University of Oregon, 1963.

Osgood, C. E. The nature and measurement of meaning. Psychological Bulletin, May, 1952, 49.

Pennock, C. D. Selecting instructional materials to meet students' reading performance level. Alberta Journal of Educational Research, 1973, 19, No. 1.

Potter, T. C. Reading skills in young children: closure and comprehension. Unpublished doctoral dissertation, University of California at Los Angeles, 1968.

Rankin, E. F. An evaluation of the cloze procedure as a technique for measuring reading comprehension. Unpublished doctoral dissertation, University of Michigan, 1958.

Rankin, E. F. The cloze procedure--its validity and utility. Eighth Yearbook of the National Reading Conference. Milwaukee, Wisconsin: National Reading Conference, 1959.

Rankin, E. F. Cloze procedure--a survey of research. In Thurston, E. S., & Hafer, L. E. (Eds.), The Philosophical and Sociological Bases of Reading, Fourteenth Yearbook of the National Reading Conference. Milwaukee, Wisconsin: National Reading Conference, 1965.

Rankin, E. F. Research design and the cloze procedure. In Figurel, J. A. (Ed.) Vistas in Reading. Proceedings of the International Reading Association, 1966, 11.

Rankin, E. F. Grade level interpretation of cloze readability scores. Twentieth Yearbook of the National Reading Conference, 1971, 30-37.

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

Ransom, P. A study to determine reading levels of elementary children by cloze testing. Unpublished doctoral dissertation, Ball State University, 1965.

Ransom, P. Determining reading levels of elementary school children by cloze testing. IRA Conference Proceedings, Part 1, 1967.

Ruddell, R. B. An investigation of the effect of the similarity of oral and written patterns of language structure on reading comprehension. Unpublished doctoral dissertation, Indiana University, 1963.

Ruddell, R. B. The effect of oral and written patterns of language structure on reading comprehension. Reading Teacher, 1965, 18, 270.

Sauer, F. M. The determination of reading instructional level of disabled fourth grade readers utilizing cloze testing procedure. Unpublished doctoral dissertation, Oklahoma State University, 1969.

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

Schoelles, I. S. Cloze as a predictor of reading group placement. April, 1971, Ed- 053-868.

Sipay, E. R. A comparison of standardized reading achievement test scores and functional reading levels. Unpublished doctoral dissertation, University of Connecticut, 1961.

Spache, G. A new readability formula for primary-grade reading materials. Elementary School Journal, 1953, 53.

Taylor, W. L. Cloze procedure: a new tool for measuring readability. Journalism Quarterly, 1953, 60.

Taylor, W. L. Application of cloze and entroph measures to the study of contextual constraint in samples of continuous prose. Unpublished doctoral dissertation, University of Illinois, 1954.

Taylor, W. L. Cloze' readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology, 1957, 91.

Weaver, W. Theoretical aspects of the cloze procedure. In Thurston, E. S. & Hafner, L. E. (Eds.), The Philosophical and Sociological Bases of Reading, Fourteenth Yearbook of the National Reading Council. Milwaukee, Wisconsin: National Reading Conference, 1965.

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

VITA

Donna Wheeler Hicks

Candidate for the Degree of

Doctor of Education

Thesis: VALIDATION STUDY OF THE HICKS CLOZE-READING TEST FOR GRADES TWO - SIX

Major Field: Elementary Education

Biographical:

Personal Data: Born in Hominy, Oklahoma, August 26, 1935, the daughter of Mr. and Mrs. Charles E. Wheeler. Married Raymond L. Hicks, October 28, 1955, in Edmond, Oklahoma; Two daughters, Essa Gayle and Paula Jean.

Education: Attended elementary schools in Oklahoma, Kansas, and Texas; graduated from Refugio High School, Refugio, Texas, in 1953; received a Bachelor of Science degree in Business Education and Elementary Education from Central State University, Edmond, Oklahoma, in May, 1958; received a Master of Teaching Degree from the same institution in 1961; completed the requirements for the Doctor of Education degree at Oklahoma State University, Stillwater, Oklahoma, in May, 1974.

Professional Experience: Employed as an elementary teacher for thirteen years by the Oklahoma City Public Schools, Oklahoma City, Oklahoma, beginning in 1958; taught grades one through four and primary special education for one year; employed as a graduate assistant at Oklahoma State University, 1972-1973; instructor in reading improvement courses, fall, 1972; instructor in reading methods course and supervised student teachers, spring, 1973; supervisor of reading clinic, summer, 1973.

Professional Organizations: Delta Kappa Gamma, Alpha Chi, Pi Omega Pi, International Reading Association, American Association of University Women.