Cloze Procedure as a Proficiency Test of Chinese as a Second Language

by

Wang Xie
B.A., East China Normal University, 1985

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF ARTS

in the Faculty of Education

ACCEPTED
FACULTY OF GRADUATE STUDIES
[redacted]
DEAN
DATE 91/09/11

We accept this thesis as conforming to the required standard

[redacted]

Dr. W. Muir, Supervisor

[redacted]

Dr. J.O. Anderson, Departmental Member

[redacted]

Dr. L.O. Ollila, Outside Member

[redacted]

Dr. R. King, Outside Member

[redacted]

Dr. J. Esling, External Examiner

Supervisor: Dr. Walter Muir

ABSTRACT

The purpose of this study is to explore the usefulness of cloze procedure as an instrument to measure the proficiency of students learning Chinese as a second language.

Two aspects of cloze format were tested: (a) deletion method, both every fifth word deletion and random deletion and (b) textual material, both familiar and unfamiliar material.

A class of third-year university students, who were learning Chinese as a second language, provided the data for this study. There were a total of four cloze tests, and the English vocabulary sub-test from the Nelson-Denny Reading Test administered in the study.

The Kuder-Richardson 20 formula was applied to calculate the test reliability. The validity of the tests was determined by correlating each of the cloze tests with the students' final exam mark and final grade.

It was found that, regardless of the deletion method, cloze tests using familiar material possessed high reliability coefficients (.94 and .85). Also, cloze tests using familiar materials yielded satisfactory validities (.89, .83, 87, and .78).

There was no significant correlation between the English vocabulary test and the students' Chinese proficiency. A questionnaire survey also revealed that the students involved in this study had a clear understanding of, and a generally positive attitude towards the cloze tests. The results of the study indicated that cloze tests, using familiar materials, were valid and reliable predictors of students' proficiency in Chinese as a second language.

Examiners

[redacted]

Dr. Walter Muir

[redacted]

Dr. John Anderson

[redacted]

Dr. Lloyd Ollila

[redacted]

Dr. Richard King

[redacted]

Dr. John Esling

TABLE OF CONTENTS

LIST OF TABLES

List of Figures

Acknowledgements

I wish to express my sincere appreciation to Dr. Walter Muir, Chairman of the Supervisory Committee, for his support, guidance, and extensive patience throughout the study as well as throughout my graduate program.

Sincere appreciation is also extended to members of my committee, Dr. Lloyd Ollila and Dr. John Anderson, for their assistance and support.

I also want to thank Dr. Richard King, not only for his participation as an outside member of the committee, but also for his valuable support and guidance during my preparation and administration of the cloze tests.

Further appreciation goes to Dr. Brian Harvey, for his concern and encouragement throughout my degree.

As well, I am especially grateful to the students in Chinese 300 who volunteered their time as participants in the study.

Finally, I would like to acknowledge and express my love and appreciation for my wife, my grandfather and my parents, who were understanding and supportive throughout the whole period of my study in Canada.

Chapter 1

INTRODUCTION

Since its creation in the 1950's by Taylor (1953), cloze procedure has been widely used in testing English as a second language (ESL) (Soudek & Soudek, 1983). Although its original use was to be a testing device of the readability of English reading materials, it has since been developed and used as a reliable, valid, and efficient instrument for measuring reading comprehension and for estimating the level of proficiency of ESL learners (Darnell, 1968).

Although the the cloze test has been widely researched with regard to ESL proficiency (Anderson, 1983; Oller & Conrad, 1971; Spolsky, 1969), the literature contains relatively few reports of attempts to measure proficiency in languages other than English. However, in a few studies cloze procedure has been used to measure language proficiency in Japanese, Korean (Anderson, 1976) and other languages (Briere, Clausing & Pincell, 1978). Only one report has been found, however, in which cloze procedure was used as a measuring device for reading comprehension in Chinese as a first language (Anderson, 1976). When conducting his experimental study in Singapore, Anderson (1976, p.112)

claimed that "No previous studies have been reported where cloze procedure has been applied to the Chinese language." No similar studies have been found since. The use of the cloze procedure as a formal proficiency test of Chinese as a second language (CSL) has not appeared in the literature.

In view of the fact that one fourth of the world's population speaks Chinese, and that more and more western people are learning CSL, the issue of developing an efficient testing device for Chinese proficiency has been raised (Liu, 1989). This offers an interesting topic for the researchers working in the field of language test development.

The Problem

Modern Chinese language testing has been greatly influenced by the ancient tests of the Imperial Examination for selecting mandarins or civil servants. The Imperial Examination system lasted almost fifteen hundred years and was based on a stereotyped form of long essays. Essay writing, therefore, has played a dominant role in determining students' overall language proficiency in Chinese. Even in western countries, where research on language testing (English in particular) has been conducted both intensively and extensively, testing procedures for CSL seems not to have

been researched. To the author's knowledge, translation still plays a major role in testing CSL today. However, the subjective nature of essay and translation tests often possesses poor reliability hence jeopardizes the validity of the test. Therefore, the use of such formats has been criticized on their subjectivity (Gronlund, 1990). Since the early 1980's, some research has been done in Chinese language testing and "new" formats have been introduced, such as multiple choice, and short answer. However, cloze procedure seems to be overlooked. Under the circumstances that the cloze procedure has been successfully used in testing ESL, the benefits of using cloze procedure as a Chinese proficiency test seem worthy of study.

The problem addressed here then, is an evaluation of the effectiveness of using cloze procedure in testing proficiency of students of Chinese as a second language.

The Purpose of the Study

In response to the problem, the present study was designed to evaluate the cloze procedure as a measuring device for the proficiency of students learning Chinese as a second language. It has been assumed that cloze procedure could be equally as valid and reliable in testing Chinese

competency as it has been in testing English, although Chinese is a language which is different from English, both in syntax and the symbols used to represent it. The ultimate purpose of the study, therefore, was to evaluate the cloze procedure used as a Chinese proficiency test for CSL teachers. The study had two aspects: first, to develop a set of cloze tests in Chinese and, second, to evaluate the tests.

Chapter 2

LITERATURE REVIEW

Background to the Use of Cloze Procedure

The cloze test has become one of the most frequently used tests of language achievement since its development in the 1950's (Soudek & Soudek, 1983). The procedure was proposed by Taylor (1953) to determine the difficulty of reading materials for a native speaker of English. Since its creation, several varieties of the cloze procedure have been used to measure reading comprehension, language proficiency, knowledge of vocabulary, the relative difficulty of passages of prose, and even the reader's intellectual capacity as well (Soudek & Soudek, 1983).

Although originally concerned with native speakers of English, the cloze procedure has been frequently considered as a useful measuring technique in ESL learning situations. Anderson (1971) used cloze procedure to determine the readability and appropriateness of English passages for an ESL class or individual. Jongsma (1971) reported that cloze procedure was a valid and reliable measure of both specific and general reading comprehension for ESL students. Since

the 1970's, considerable interest has been shown in the use of cloze tests as measures of overall language proficiency in ESL situations (Aitken, 1977; Anderson, 1883; Oller & Conrad, 1971). Alderson (1978, 1979a, 1979b, 1979c) has reported on the relationship between cloze tests and ESL proficiency, the use of different scoring procedures on cloze tests and the effect on cloze tests of varying the frequency with which words are deleted from text when using the pseudo-radom deletion procedure.

The Theory of Cloze Procedure

The cloze procedure is usually explained in two aspects, one of which is the "law of closure" from Gestalt psychology, the other being the redundancy in human language.

Cloze and Closure

The traditional explanation of the dynamics of cloze procedure centered around the Gestalt concept of closure (Taylor, 1953).

Gestalt theory was developed by a school of thought in Germany and was first published by Wertheimer and others in 1912 (Hill, 1971). Gestalt theory views concrete phenomena,

such as visual figures, as well as abstract phenomena, such as memory or learning, as organized structures or wholes, rather than as combinations of separate units. The focus is on Ghanzheit, which could be translated as wholeness, or completeness.

Gestalt suggests that the human perception of form comes from the relationships among the parts of the form. In this process of perception, the parts might lose their former properties and take on new properties which are determined by the form of the whole pattern. For example, each tone comprising a musical melody played in isolation has a certain subjective pitch and loudness, but in the context of a melody the subjective pitch or loudness of the tone depends on the relation it has to its neighbors (Hill, 1971).

Within this theory, the law of closure is of vital importance. It reflects the tendency of human beings to perceive unfinished or incomplete figures as completed entities, to fill in the gaps in broken patterns. Therefore, human beings have the ability to complete a broken piece of music or some visual patterns.

Gestaltists believe that learning follows a sequence through which one first understands the whole of broken issues, and then grasps the individual details. Similarly,

the cloze test requires the subjects to complete the whole by filling in the missing words.

Based on these concepts, Taylor (1953) coined the word "cloze" which in spelling, though not in pronunciation, is an intentional corruption of the verb "to close", and is intended as a reminder of the Gestalt "law of closure." In the cloze procedure selected words are deleted from prose text and replaced by blanks. Filling in the blanks by "guessing" the missing words is, according to Taylor's notion, a special kind of closure and hence the term "cloze."

When cloze procedure was first reported as a potentially useful practice, Taylor explained it in terms of the tendency of the human mind to complete patterns. He thought of cloze procedure as a means of measurement, and described the cloze unit (that is, the deleted word) as follows (Taylor, 1953, p.145):

> At the heart of the procedure is a functional unit of measurement tentatively dubbed a "cloze." It is pronounced like the verb "close" and is derived from "closure." The last term is one Gestalt psychology applied to the human tendency to complete a familiar but not-quite-finished pattern - to "see" broken circle as a whole, for example, by mentally closing up the gaps ... One can complete the broken circle because its shape or pattern is so familiar that, although much of it is

actually missing, it can be recognized any way. The same principle applied to language ...

Taylor's explanation has resulted in many controversies among researchers. Louthan (1965) agreed that cloze procedure could be explained by the Gestalt concept. Oller and Conrad (1971) supported this by saying that the restoration of words deleted from a selection of prose in order for the passage to make sense was a special use of the ability to complete broken patterns.

Weaver (1965) and Anderson (1971) however, did not agree that cloze procedure was an aspect of closure. Weaver argued that cloze responses were cognitive rather than perceptual and that the cloze task was subjective rather than objective. He criticized the Gestalt view of cloze by suggesting cloze was a type of language search process rather than the mere completion of a perceptual task. Anderson supported this by saying that cloze procedure required subjective efforts and that it did not make use of objective information in the same way that fill-in-the-blanks or sentence completion tests did. Then both of them argued that cloze procedure measured comprehension not "closure."

Rye (1982) further claimed that the theoretical model for cloze procedure based on Gestalt psychology had "serious weaknesses." He indicated that this model might describe

what happens in some cases of perception, but filling in cloze deletion is not about "seeing" patterns in the sense of seeing a visual pattern. He went on to explain his own theory by saying that cloze procedure was essentially a cognitive task and the reader had to reason and construct suggestions to fill the gap on the grammatical "pattern" in language. The reader's innate ability to produce grammatically appropriate sentences would help determine the grammatical class of the word to be produced. However, the completion of meaning, based on understanding and reasoning, is a cognitive task. Rye (1982) also suggested that the term "construction procedure" might be a more accurate title for the process. Ohnmacht, Weaver and Kohler (1970) concluded that there was no strong relationship between the gestalt principle of closure, and the completion of items in a cloze passage.

It seems that the matter of whether or not cloze is related to closure appeared to depend largely on an individual's interpretation of the Gestalt definition. It might be more productive to try other explanations of the cloze procedure by referring to other learning theories and examining the nature of human language learning itself.

Cloze and Redundancy

In order to study further the cloze procedure from the perspective of human perception, an important problem must be solved. That is, whether features are successively identified one by one, or whether some or all of the features are identified simultaneously. Cattell (1880), cited in Travers (1982) studied this problem a hundred years ago and he concluded that the visual system could process several pieces of information simultaneously.

Travers (1982) claimed that in order to read a word in context, one may not have to recognize more than one or two letters in order to identify the word with considerable precision. He went on by saying, "that is why proofreaders often do not notice errors in printing. They read what they expect to see rather than what is there." (Travers, 1982, p.72) Reading precisely what is on the page is a much slower process than reading in terms of expectation and conjecture.

Travers (1982) claimed, further, that what one has just read usually allows one to anticipate what one is to read next; that is, printed material includes considerable redundancy, meaning that there is an overlap in the information provided by one segment of the material and other segments of the same body of material. Without such

redundancy, and the anticipation it provides, reading would be a much slower process. This notion is supported by Spolsky (1971) who claimed that oral messages can be understood even though a good proportion of detail is omitted or masked; in other words, every message contains elements that can be omitted without leading to a breakdown in communication.

Manis and Dawes (1961) suggested that language redundancies generally have been thought to provide safeguards against errors in communication. Because of such redundancies, a message blocked at one part is not necessarily blocked at all parts. Due to the redundancies inherent in the language, parts of sequences can be anticipated. A person reading or listening with understanding constantly makes tacit predictions of what will come next.

Thus, from the foregoing, it could be deduced that the cloze procedure, that is, the family of techniques for systematically deleted portions of text, is a method for testing the learner's internal system of what Taylor(1953) referred as "grammatical expectation." To illustrate this idea, Taylor (1953, p.419) cited the following argument which he attributed to Charles E. Osgood:

> Some words are more likely than others to appear in certain patterns of sequences. "Merry Christmas" is a more probable combination than "Merry birthday." "Please pass the__." is more often completed by "salt" than by "sodium chloride" or "blowtorch." Some transitions from one word to the next are, therefore, more probable than others.

According to Taylor (1953, p.418), Osgood argued that the foregoing was a product of the redundancy of natural language:

> "Man coming" means the same as "A man is coming this way now." The latter, which is more like ordinary English, is redundant; it indicates the singular number of the subject three times(by "a", "man", "is"), the present tense twice("is coming" and "now"), and the direction of action twice("coming" and "this way"). Such repetitions of meaning, such internal ties between words make it possible to replace "is", "this","way", or "now" should any one of them be missed.

Cloze procedure, then, is possible because of the inherent redundancy in the English language. A cloze score, therefore, indicates how well the message of the sender is understood by the receiver.

Characteristics of the Cloze Procedure

Oller (1979, p.341) compared the ability of a perceiver to complete broken patterns with the ability of a reader to complete the following mutilated portions of text:

1) one, t__, t____, f___, _ive, __x, ____n, ...
2) Four _____ and seven _____ ago _____ _____ ...
3) After the mad dog had bitten several people he was finally s_ght_d n__r th_ _dg_ _f t_wn _nd sh_t b_ a local farmer.
4) It is true that persons _____ view the treatment of mental _____ from a clinical perspective tend _____ explain socioeconomic and ethnic differences _____ biological terms.

In Example 1 we have no difficulty in supplying the missing letters of the words "two" and so on. The series is so highly redundant that we can anticipate the series on the basis of very little textual information. Example 2 is also highly redundant if one happens to know the first line of Lincoln's Gettysburg address, otherwise it is more difficult. In any case, even when the original text is not stored in its entirety or at least in some recoverable form in memory, missing or mutilated portions may nonetheless be recoverable by a creative process of construction as illustrated by Examples 3 and 4. The mutilated words in 3 are "sighted near

the edge of town and shot by" and in 4 the missing words are "who", "retardation", "to" and "in" (Oller, 1979).

These examples provide a basis for the comparison of the notion of pattern completion in Gestalt psychology and the concept of closure in relation to the processing of text. When the material is almost completely redundant, (e.g., filling in the missing letters in the series of words in Example 1, or filling in the missing words in a text that has been committed to memory), the task would seem to be somewhat like the process of filling in the gaps in imperfect visual patterns. However when the material is not familiar and is therefore less redundant, the power of the generative mechanisms necessary to fill in the gaps or to restore distorted portions of text would seem to be more complex than the simple visual cases.

Anderson (1983) reported that the difficulty of a cloze test is highly related to the nature of the passage selected and the frequency of the word deletion. The more sophisticated the passage and the more frequent the deletion (that is, the more broken the passage), the more difficult the text becomes. Also, according to Oller (1979), three types of knowledge are called for in order to complete a cloze passage correctly - linguistic knowledge, textual knowledge, and knowledge of the world.

Similarly, in visual completion, if the figure field becomes broken so that the point of probable fixation is shifted to the ground field, it is harder to see the picture. The point of fixation is determined by the arrangement of the lines and by the relative size of the interspersed ground. Thus, the student's ability to supply the missing parts and to perceive in his or her own mind the figure in its entirety rests partly upon these external conditions and partly upon the maturity and social experience of the observer (Hartmann, 1935).

McKenna and Robinson (1971) indicated that there is a certain similarity between Gestalt psychologists' law of closure and Guthrie's primary law of conditioning (Guthrie, 1952). Thus, in terms of learning theory, cloze procedure is somewhat like Guthrie's maze procedure, which presents the reader not with blanks but with irrelevant words periodically embedded in the text. It therefore constitutes a multiple-choice cloze format in which the student must continually remove the unnecessary elements in order to choose the one sensible path through a verbal "maze."

Actually, cloze and maze are logical opposites. In cloze, verbal information is removed from the passage; in maze, it is added. Thus, while cloze is a task of recall, maze is one of recognition. This relationship makes maze a particularly valuable process for better understanding cloze,

and also increases its own value in both assessment and instruction.

Reliability and Validity of the Cloze Test

Reliability is an index of the extent to which a test is consistent or stable. This test characteristic can be estimated through test-retest, equivalent-forms, split-half or Kuder-Richardson procedures (Borg & Gall, 1989).

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

The reliability of cloze tests when administered to ESL students has been demonstrated in numerous studies. Brown (1980) reported reliability coefficients that ranged from .61 to .95 for various scoring methods, which suggested cloze tests are generally reliable (see Table 1). The differences of the reliability in all these studies could be a result of different sample sizes, number of items and scoring methods as well.

Table 1
Language Proficiency Cloze Reliability Studies. (Brown (1980)

Study	Sample Size	Sample Level	Number of Items	Scoring Method*	Reliabi. Coeffi.
Darnell (1970)	48	Univ. (ESL)	50	CLZNT	.86
Oller (1972)	398	Univ. (ESL)	50	EX AC	.80-.92 .90-.95
Pike (1973)	430	Univ. (ESL)	25	EX CLZNT	.78-.91 .83-.85
Wijnstra & Wageningen(1974)	100	Grades 4-6	45	MC	.88
Jonz (1975)	125	Grade 7-12 (ESL)	65	MC	.95
Hinofotis (1976)	170	Univ. (ESL)	50	EX AC	.61 .85
Jonz (1976)	33	Univ. (ESL)	33	MC	.76

*CLZNT = clozentropy; EX = exact; AC = acceptable;
MC = multiple-choice.

Validity is an estimate of how well a test measures what it claims to be measuring. The validity of a test is usually determined through content validity (how well the test samples the content of the subject-matter), construct validity (how test performance can be described psychologically), or criterion-related validity (how well the "new" test correlates with the criterion measure, such as another test which measures the same subject matter).

Bormuth (1968) claimed that cloze procedure had content validity. Since a cloze test was constructed on the material a student would be required to read, it obviously provided a demonstration of his/her achievement of that material.

However, criterion-related validity is also used because it is the easiest to document and defend: unlike other forms of validity, it is expressed in a correlation coefficient. Darnell (1968), using a cloze test and modified scoring system on a group of ESL students, obtained a correlation of .83 with the test of English as a Foreign language (TOEFL) total test score. Oller (1972) using two different scoring methods, but the same cloze test, obtained correlations of .75 and .83 with the University of California at Los Angeles ESL Placement Examination. Stubbs and Tucker (1974) also reported a correlation of .76 between a cloze test and the English Entrance Examination of the American University of

Table 2
Language Proficiency Cloze Validity Studies. (Brown,1980)

Study	Sample Size	Sample Level	Criterion	Scoring Method*	Validity
Conrad (1970)	35	Univ. (ESL)	ESLPE,UCLA Form 2b	EX	.88
Darnell (1970)	48	Univ. (ESL)	TOEFL	CLZNT	.84
Oller (1972)	100	Univ. (ESL)	ESLPE,UCLA	EX	.80-.89
Oller (1972)	398	Univ. (ESL)	ESLPE,UCLA	EX AC	.73-.87 .80-.89
Irvine et al (1974)	159	Univ. (ESL)	TOEFL	EX AC	.78 .79
Stubbs &Tucker (1974)	155	Univ. (ESL)	Eng.Entr.Ex. Amer. Univ. of Beirut	EX AC	.71 .76
Hinofotis (1976)	107	Univ. (ESL)	Placement T. CESL, SIU	EX AC	.80 .84
Hinofotis (1976)	52	Univ. (ESL)	TOEFL	EX AC	.71 .79
Hinofotis & Snow (1980)	66	Univ. (ESL)	Placement T. CESL, SIU	EX AC MC	.71 .74 .63

*EX = exact; AC = acceptable; CLZNT = clozentropy;
MC = multiple-choice.

Beirut. The validity coefficients from a variety of studies are also presented in Table 2. The range of coefficients reported between cloze tests and criterion measures is from .63 to .89, thus cloze procedure seems to provide fairly valid measures of overall ESL proficiency.

The cloze validity and reliability have been investigated throughout the 1980's. DeSanti and Sullivan (1985) studied cloze reliability by testing a total of 456 subjects representing grades 3, 5, 7, 9, and 11 with 200 passages. They found 100 reliability coefficients which ranged from .55 to .96. DeSanti (1989) studied concurrent and predictive validity using a total of 408 subjects representing grades 4, 6, 8, 10, and 12 of a large suburban public school system. Each participant completed three alternative cloze passages of fifty items in each. Pearson product moment coefficients between the cloze procedure scores and the subjects' Comprehensive Test of Basic Skills (CTBS) scores were calculated to assess concurrent and predictive validity. Of the 324 correlations analyzed (ranging from .27 to .76), 314 (97%) were statistically significant.

Cloze Deletion Pattern

According to Syme (1989), various deletion patterns have been the subject of considerable research. Some research

suggested that deletion patterns either inhibit or enhance the reader's ability to use context clues to predict the missing words (Syme, 1989).

MacGinitie (1961) investigated various deletion patterns and found that deletion patterns with less than four words between deletions did not provide the reader with enough context. Culhane (1970) suggested every tenth word deletion for textual materials that are "fact laden", and every fifth word deletion for narrative materials. However, the rationale behind this needs further investigation.

Helfeldt and Henk (1985) compared a totally random format and a traditional every fifth word deletion pattern using three reading passages. A number of sequences were employed and administered along with the Nelson-Denny Reading Test (NDRT) and Iowa Silent Reading Test (ISRT). The subjects were college students. They found that both random and fifth word deletion were significantly correlated with NDRT ($r=.60$, $p<.01$ & $r=.47$, $p<.02$) and ISRT ($r=.71$, $p<.01$ & $r=.62$ $p<.01$). Total random deletion correlations were higher than every fifth deletion correlations, but these differences were not significant. Higher coefficients of reliability were found for the total random test, but statistically significant differences were not found. In another study using 64 sixth grade pupils, Henk and Helfeldt (1985) found no significant differences between every fifth word deletion and total random deletion patterns.

Cloze Test and Related Course Achievement

Several studies have used cloze procedure to predict related students' course achievement. Van Every (1981) used 223 students enrolled in the course over two academic years and gave them a 50-item cloze test developed from each of two required tests. Pearson Product Moment coefficients were then calculated between mean score in the cloze tests and the grade earned in the course. The course was structured on a mastery basis and consisted of three units, each requiring a student to take three exams to meet a minimum criterion level of 75% for a "C" in the course. Students who were able to meet the minimum criterion level in only three exams had an average of 59.5% correct on the cloze; students who needed nine exams had 45.5% average correct fill-ins. Students not passing the course had significantly lower scores than did students who earned a passing grade. Obviously, students who had a higher course achievement tended to perform better on the cloze test.

Stephens and Weaver (1986) investigated the cloze procedure as a predictor of success in college classes. Six cloze tests developed from political science textbooks were administered to students in an introductory political science course. Pearson correlation coefficients between the students' performance on cloze tests and their class grades ranged from .31 to .59. The experiment was replicated with

95 students enrolled in introductory classes in three other disciplines: sociology, psychology, and biology. In each case, cloze passages were constructed using the assigned text. The overall coefficient between cloze results and final class grade was .49. When coefficients of correlation were computed by class, sociology had the highest coefficient (r=.58), and biology the lowest (r=.45).

More recently, Chance (1988) examined the relation between pupil performance on cloze tests constructed on the textbook with performance on teacher-made tests concerning the same subject matter. Twenty-five eighth-graders were administered three measures for each of five chapters of an American history textbook. First, they took a cloze pretest based on a 300-word section from the beginning of each chapter. Second, they completed a teacher-made test consisting of multiple-choice and fill-in-the-blank items on the content of each chapter. Third, they took a cloze posttest based on a 300-word passage from the end of the chapter. No significant coefficient of correlation was found between differences in both pretest and posttest cloze scores and scores on the teacher-made test. However, posttest cloze scores were significantly higher than pretest cloze scores on four of the five chapters.

An Experiment of cloze in Chinese Language

In his book "Psycholinguistic Experiments in Foreign Language Testing", Anderson (1976) reported on his interesting experiment with cloze tests constructed in Chinese.

This experiment was conducted in the island Republic of Singapore by personnel in the Research Unit of the Singapore Ministry of Education. The purpose of the experiment was to explore the possibility of the further application of cloze procedure, since no previous studies had been reported where cloze procedure had been applied to the Chinese language.

The first step of the experiment was to select five passages in Chinese which were then ranked by specialists in order of reading difficulty. The five passages were adopted as the criteria of reading difficulty. For all the passages selected an every fifth word deletion system was used.

Nine Chinese schools were then randomly-selected from all the co-educational primary schools in which Chinese was used as the official language for teaching. The nine schools were randomly assigned to three groups - Primary 4, Primary 5, and Primary 6. From each school one class was randomly selected from among the various classes in the level to which the school had been assigned. Hence the sample consisted of

three classes of Primary 4 pupils, three classes of Primary 5 pupils, and three classes of Primary six pupils, all selected at random.

To test reliability one class from each of Primary four, five, and six was randomly selected from the classes, and the test was re-administered exactly a week later. Test-retest correlations were then calculated. As a check of the validity of cloze procedure as a measure of reading ability, final term examination results in various school subjects were correlated with cloze comprehension scores for each pupil. The results suggested that the cloze ranking of the five passages for each group level might provide more accurate ordering than the specialists'. Test-retest reliability coefficients ranged from .82 to .92 for the passages separately and rose to .96 for the passages combined.

Anderson (1976) also claimed that, "when school examination results were correlated with cloze scores, Chinese correlated most highly with the cloze test, thus providing some evidence of the validity of cloze procedure with Chinese" (p.119). Unfortunately, however, there was no specific data provided to support this assertion.

It was concluded by Anderson (1976) that, 1) cloze procedure ranked passage difficulty reliably, and was in good

agreement with experts' rankings, 2) cloze procedure discriminated between pupils in different levels of Chinese competency, 3) cloze scores correlated well with Chinese and other school subjects, and 4) it made little difference to pupils' ranks whether scoring was based strictly on the original characters used in the passage (exact-word method) or on characters that made sense within the context (acceptable-word method).

In the recent literature, there were a few reports on cloze procedure used in languages other than English, such as, Chavez-Oller, Weaver and Oller (1985) dealt with Japanese and Kamil, Smith-Burke & Rodriguez-Brown (1986) with Spanish. However, as far as Chinese is concerned, nothing has been found.

Summary

Although cloze procedure was originally used for native speakers of English, it is rapidly growing in popularity in both English as a Second Language (ESL) and foreign language programs because it appears to be an effective, economical method of measuring overall language proficiency. A number of studies have reported that cloze procedure enjoys high validity and reliability. Studies also show that the

traditional every 5th word deletion pattern is quite adequate in terms of general utility and that it is the same as the total random deletion pattern. Research on the cloze test and related course achievement indicated that cloze tests generally correlated significantly with student achievement on courses of the same subject matter. The only experiment with Chinese found in the literature suggested that the cloze procedure might be applied to test Chinese language proficiency with some degree of validity and reliability.

Chapter 3

RESEARCH OVERVIEW

Rationale for the Study

Taylor (1954) suggested in his doctoral dissertation that the rationale for the cloze procedure is based on the Gestalt principle of closure. This principle states that there is a tendency for individuals to close incomplete forms so as to constitute wholes. The familiar example is the incomplete circle which most subjects see without any breaks or discontinuities. The same principle, applied to language, involves the deletion of language elements to be reconstructed by observers. The restoration of words deleted from a selection of prose in order for the passage to make sense is a special use of this ability to complete broken patterns.

It has already been demonstrated that the cloze procedure is a very effective instrument to test subjects' general language proficiency in English, both as a native and second language (Aitken, 1977). Although Chinese language is different from English, both share some common linguistic

features. In addition, due to the ideographic nature of the Chinese script, it could be assumed that the Gestalt closure principle might fit even better in cloze tests when applied to the Chinese language.

It might, therefore, be hypothesized that the cloze test could prove to be a highly efficient test of proficiency for Chinese as a second language.

Definition of Terms

Cloze procedure has developed its own terminology, and to avoid confusion the more important terms and their usages are defined below.

Cloze procedure. The cloze procedure consists of a set of rules for constructing cloze tests from samples of written materials, administering these tests to subjects and scoring them, and determining from the cloze scores the degree of readability of the written materials. The rules specify the deletion system, the length of blank, and the scoring methods.

Cloze test. A cloze test is the application of the cloze procedure to a passage of text. A predetermined number of words is deleted from a passage, and the respondent is asked

to supply the missing words. The resulting cloze score is considered to be a measure of the subject's command of the language.

Blanks. Blanks are used to denote the deletion of words. They are usually underlined and are ten spaces in length for a passage in English. In the cloze tests constructed for this study, each blank allows a space for two Chinese characters.

Deletion system. Deletion system refers to the mechanical procedure of deleting words from a passage. An any-word deletion system is the deletion of words without regard to grammatical class. A random deletion system is the deletion of words by reference to a table of random numbers. An nth deletion system is the systematic deletion of words through a passage, for example, every fifth word or every eighth word.

Scoring methods. Four scoring methods are used in cloze procedure. Beginning with Taylor (1953), both the "exact-word" (EX) and "acceptable-word" (AC) scoring methods have been employed. The EX method requires the complete restoration of only the original word omitted, while the AC method usually counts any contextually acceptable answer as correct. With native speakers, there seems to be little difference in reliability and validity between these two scoring methods (Taylor, 1953). In most teaching situations

where cloze is used as an exercise the AC approach seems to be more appropriate for it does not penalize creativity and can be used as a starting point for useful and interesting class discussions.

A further refinement of this idea is the clozentropy (CLZNT) scoring method, which gives weighted credit for responses that are the same as those given by a native speaker of the language during a pretest (Darnell, 1970).

One final scoring method, multiple-choice scoring (MC), has been investigated by Jonz (1976). With this technique, students are provided alternative answers from which to choose the correct answer for each blank.

Research Questions

The purpose of the present study is to evaluate the cloze procedure as a Chinese proficiency test for learners of Chinese as a second language. Therefore, the research questions focus on the validity and reliability of the cloze tests and related issues:

Question 1 Can the cloze procedure validly test students' proficiency in Chinese as a second language?

Question 2 Can the cloze procedure reliably test students' proficiency in Chinese as a second language?

Question 3 Is there a relationship between the results of cloze tests using familiar and unfamiliar materials?

Question 4 Is there a relationship between random deletion and every 5th word deletion method?

Question 5 Is cloze procedure a better predictor of the students' Chinese proficiency than an English vocabulary reading level test?

Chapter 4

METHOD

Subjects

The subjects were 11 third-year students enrolled in the Chinese language program of the Department of Pacific and Asian Studies, University of Victoria. However, one subject was unable to complete all of the four cloze tests and was deleted from the sample. Thus, the study was based upon the ten students who completed all the cloze tests.

Instrumentation

Following the English cloze test format, the author constructed a set of cloze tests in Chinese. The test format was produced in clear print handwriting of the author. In order to examine the different testing materials and different deletion methods, five cloze tests (including a practice test), were designed (see Appendix A).

Test Material

The materials used for the tests came from the same source, which was entitled "Cheng Yu Gu Shi Xuan" (Beijing Language Institute, Beijing Foreign Languages Publishing House, 1979), that is, "Annotated Chinese Proverbs." These were all short parallel passages in simple narrative style, with each passage telling a story about a famous Chinese proverb that derived from a well-known ancient Chinese story. The reason for selecting such passages was to minimize the differences among the test passages, so that the test materials could be selected as consistent to each other in terms of content and style as possible.

To construct the cloze tests, the author decided to expose to the students both materials they were familiar and unfamiliar with. Therefore, two familiar materials were text selected directly from the students' textbook, which had included seven "annotated Chinese proverbs" from "Ceng Yu Gu Shi Xuan." The textbook was compiled by the instructor for use in the course. The text selected for the cloze had already been studied by the students. It was assumed that this method would make the cloze procedure a reasonable task for the students of a lower level of Chinese proficiency.

The unfamiliar passages were selected by the author. They were also two proverbial stories from "Cheng Yu Gu Shi

Xuan" other than those in the textbook. It was confirmed by the instructor that these two passages had not been studied by the students. Since the materials were unfamiliar to the subjects, they were assured by the instructor that they had been carefully chosen at an appropriate level, which was identical in terms of reading difficulty to that of the familiar passages so that the students should be able to understand them.

Test Length

Regarding test length, Taylor (1956) suggested that fifty blanks were usually required to yield a stable estimate of individual performance. However, Oller (1979, p.365) argued that, "there may be special circumstances where this suggestion should be disregarded. For instance, it would make no sense to avoid using a short text when it is that same short text that has been studied." The use of short texts in this study was decided on the basis of: 1) students' reading capacity of Chinese could not be expected to be too high, 2) time available for the tests was limited, and 3) the proverbial stories in the textbook were all short. Therefore, all the passages selected for this study were approximately 130 Chinese characters in length, which allowed 25 blanks. It was hoped that this modification be generally in accordance with the students' reading capability in terms of the passage length.

Deletion Technique

As for the cloze deletion technique, two methods were applied: (a) the first method was a random deletion by which approximately one fifth of the words (characters) were randomly deleted from the passage (consecutive deletions were avoided), and (b) the second method was a fixed-ratio deletion approach by which every fifth word (character) was deleted from the passage.

The two deletion methods were applied both to familiar and unfamiliar materials to provide a total of four test formats for the study. They were: test using familiar material with every fifth word deletion; unfamiliar material with every fifth word deletion; familiar material with random deletion; and unfamiliar material with random deletion.

Scoring Method

For each type of the cloze format one testing session was provided. Thus, there were a total of 4 formal test sessions for the whole study.

All of the responses were scored by the author using both 'exact-word' method and 'acceptable-word' method (Oller, 1979; Taylor, 1956). Miller and Coleman (1967) reported a correlation of .99 between the two deletion methods.

Procedure

Before the actual testing began, the author sat in the students' classroom, with permission of the instructor, for several lessons in order to become familiar with the students and the content of the course so that maximum cooperation could be achieved.

Prior to the formal testing, a practice cloze test with the same characteristics as the formal tests was administered to help the subjects become familiar with the task. The practice test was administered on November 2, 1990. During this test, the author tried to create a casual and relaxed atmosphere for the students and answered every question raised. This seemed to be very successful as the students began to understand that this was not a "test" but rather an exercise that could help them learn Chinese.

To assure complete understanding, the author announced the correct answers to all items immediately after all the students had finished the practice test. A short period of discussion followed to help the students fully understand the linguistic issues and the cloze task.

The formal cloze tests started exactly one week after the practice test. They were administered over four consecutive weeks beginning from November 9, and ending on

November 30, 1990. Each of the four formal tests was administered after a regular Chinese class on four successive Friday mornings. Following a usual group testing procedure, the author supervised in the testing room to ensure that all the students worked independently within a time limit of about 15 minutes, which had been set by the author according to the earlier practice test. Although the students had already taken a practice cloze test, the author still felt it was necessary to give them complete instructions at the beginning of each test:

> "The following Chinese passage has certain characters missed out. Each blank in the passage stands for just ONE character that is missing. Read the passage carefully and write in the missing character. You may find clues in the passage to help you find the missing characters. If you can think of more than one character that would fit into the gap, choose the one that you think the author would have used. If you can not think of a word, make a guess."

To avoid a learning effect on the subsequent tests, there was no discussion of the earlier test results. After each test, the author collected the test sheets and marked them later.

Over the four week period, the tests were administered in the following order: test using familiar material with every fifth word deletion; unfamiliar material with every fifth word deletion; familiar material with random deletion, and unfamiliar material with random deletion. It was

considered that this would have a minimum order effect on the test results.

On December 4, the subjects were administered The English vocabulary sub-test from the Nelson Denny Reading Test (Brown, 1973). The purpose of this was to determine the students' vocabulary reading level in English and its relation to their Chinese proficiency. Due to the time limit, the comprehension part of the test was deleted.

After the vocabulary test, the students were also asked to answer a questionnaire (see Appendix B) prepared by the author. The questionnaire was composed of four questions of multiple-choice referring to the students' understanding of the cloze test, and one short answer question asking for comments on the cloze.

Chapter 5

RESULTS AND DISCUSSION

The primary purpose of this study was to evaluate the effectiveness of the cloze procedure as a measure of Chinese proficiency in a second language learning environment. In the light of this purpose, the results are discussed to answer the research questions proposed in Chapter 3, with an emphasis on the validity and reliability of the cloze test. In addition, a summary of the questionnaire results will be provided, revealing the the subjects' point of view towards the cloze test.

Overview of the results

To obtain the results from the study, criterion-related validity was determined by correlating cloze scores with two types of scores. The first was the students' final exam scores. The second was their final grades, which combined mid-term grades, several quiz grades and classroom presentation grades as well.

The Kuder-Richardson Formula 20 (KR-20) was applied to calculate the reliability coefficient (Kuder & Richardson, 1937). For all the correlations, the Pearson product-moment correlation formula was employed.

Table 3, presents basic statistics of the four cloze tests as well as the students' final exam scores, final grades, and scores from the Nelson-Denny English vocabulary test.

All the cloze tests in the study consisted of 25 blanks, thus, the full score of each cloze test was 25. The students' final grades have been transferred into percentages based on the following scale:

A+ = 97(100-95), A = 92.5(94-90), A- = 87.5(89-85),
B+ = 82.5(84-80), B = 77.5(79-75), B- = 72.5(74-70),
C+ = 67.5(69-65), C = 62.5(64-60), D = 54.5(59-55)

Therefore, the mean scores of the four cloze tests indicated, that the difficulty of the tests are at the appropriate level for the subjects concerned. This is supported by Bormuth (1968) who compared multiple choice comprehension scores to cloze test scores based on the same material. He found that a 44% cloze score corresponded to a 75% comprehension score and a 57% cloze score corresponded to a 90% comprehension score. None of the mean scores in the present study is below 44% which was considered a frustration level (Rankin, 1970).

Table 3
Descriptive Statistics

Test	Sample Size	Mean Score	Std. Dev.	Minimum Scores	Maximum Scores
Familiar every 5th deletion	10	15.10	6.64	2	23
Unfamiliar every 5th deletion	10	11.80	3.65	4	18
Familiar random deletion	10	14.60	4.12	10	22
Unfamiliar random deletion	10	12.40	2.91	8	17
Final exam	10	82.86	6.58	75	95
Final grade	10	76.80	11.56	57	94
Vocabulary test	10	51.60	10.92	24	62

The maximum score on the vocabulary test was 62, while the minimum score was 24. This wide range (38 points) indicated that the students' English vocabulary reading ability was variable, as far as their native language is concerned. Therefore, the subjects in this study could be considered as representative of normal college students related to test norms.

It is noticeable that the tests using familiar materials had a higher mean than the tests using unfamiliar materials. For the tests using familiar material, every fifth deletion test had a mean of 15.10 and random deletion 14.60. For unfamiliar material tests, every fifth deletion had a mean of 11.80 and random deletion 12.40. This indicated that the familiar materials were easier for the students than the unfamiliar ones. The tests using familiar material also possessed higher standard deviations (6.64, 4.12 in familiar material tests and 3.65, 2.91 in unfamiliar material tests), which reflected a considerable range of scores in those tests. That is to say, the majority of scores in familiar every fifth deletion test were between 15.10 + 6.64 and 15.10 - 6.64, while the majority of scores in unfamiliar random deletion test were between 12.40 + 2.91 and 12.40 - 2.91. These results were further demonstrated by score distributions in Figure 1.

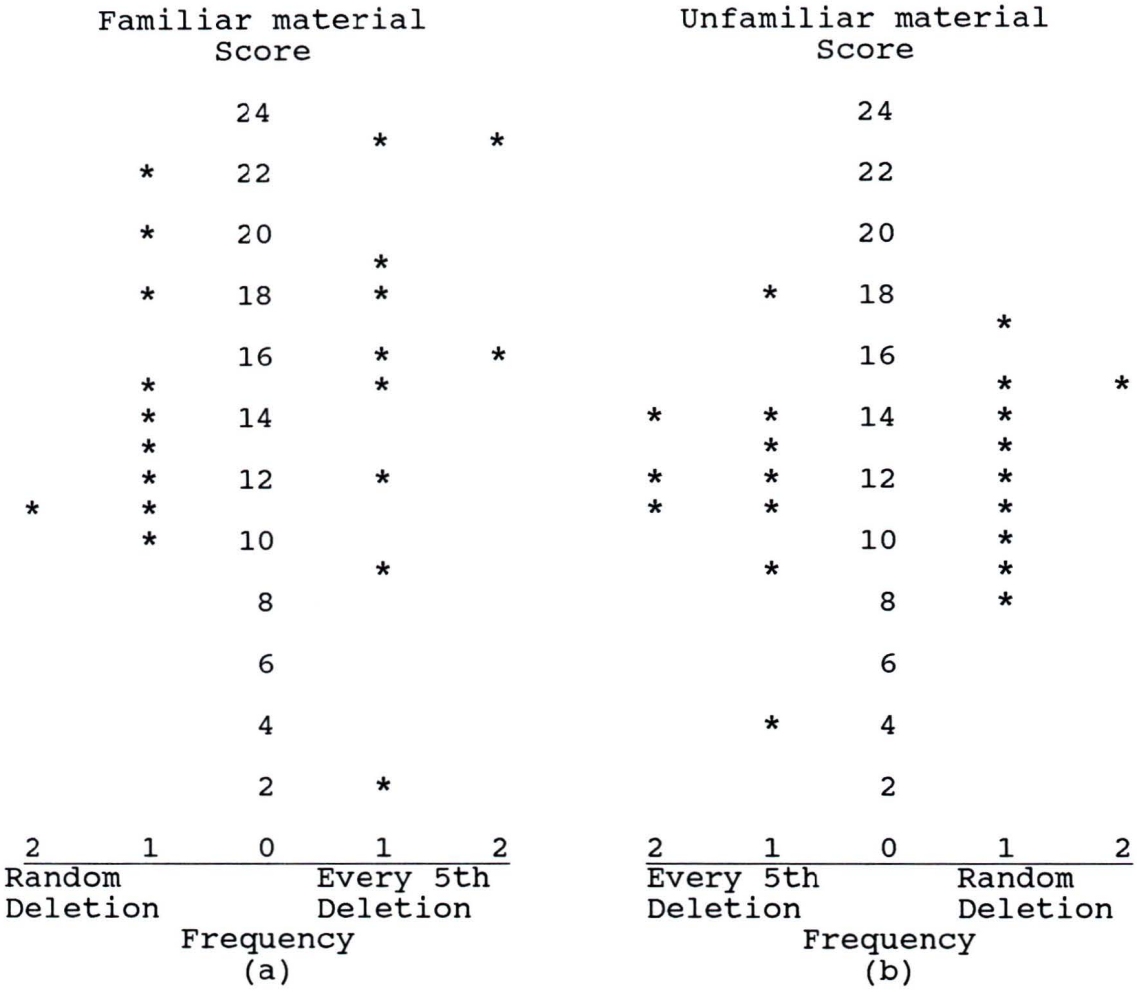

Figure 1

Cloze Test Score Distributions

For the tests using familiar material, every fifth word deletion score were evenly distributed in a range between 2 and 23, with two scores on 16 and 23. The curve looks like a negatively "skewed" distribution. For random deletion, scores evenly distributed between 10 and 22, with only two scores on 11 (Figure 1 a). For the unfamiliar material, every fifth word deletion score distribution appeared to be "normal", and random deletion scores were evenly distributed between 8 and 17.

There was no regular pattern among these distributions, but the difference was that, scores of tests with familiar material, which had higher SD's, scattered around their means while unfamiliar material test scores, which had lower SD's, clustered together. The ranges of the two tests using familiar material were 21 and 12 points, while those of the two tests using unfamiliar material were 14 and 9. Therefore, the two tests using familiar material more successfully discriminated the "good" students from the "poor" students in terms of their Chinese proficiency.

By examining each of the cloze items (blanks), it was found that twelve items in familiar material tests had a difficulty index of 1.00 (all the students responded correctly on these items), while only four items in unfamiliar material tests were found with the same index. On the other hand, however, there were three items in unfamiliar

material tests with a difficulty of 0 (all the students failed to respond correctly), and two items in familiar material tests with the same index. Apart from the material difference which might have caused the uneven item difficulties, the common features that tended to decrease the item difficulties (up to .80) were: 1) items with high redundancy in the context, that is, characters deleted had appeared in previous content, and 2) characters used more frequently by native Chinese, such as "有一个人" and "大家." Items with lower difficulty index (down to .20) were: 1) characters possessing too many strokes, such as "最" and "穿"; 2) functional words such as "把", "得" and "地" and 3) measure words in Chinese, such as "一壶酒."

Cloze Reliability

The KR-20 formula was employed to determine the internal consistency of the scores of four cloze tests. The results are shown in Table 4.

The cloze tests using familiar materials yielded reliability coefficients of .85 and .94 which also appear to be high indexes when compared with those in the literature (ranging from .61 to .95).

Table 4
K-R20 Reliability of the Cloze Tests

Cloze	Familiar every 5th deletion	Unfamiliar every 5th deletion	Familiar random deletion	Unfamiliar random deletion
Reliability	.94	.66	.85	.49

It is noted that cloze tests using familiar materials possess both high SD's and high reliabilities. Cloze test using familiar material with every fifth word deletion had the highest SD (6.64), therefore, the highest reliability (.94); unfamiliar random deletion with the lowest SD (2.91) had the lowest reliability (.49). The high reliabilities also paved the way for high validities.

Cloze Validity

As was indicated earlier, Bormuth (1968) stated that the cloze test has "good content validity" in that it is constructed exclusively on the material a student would be expected to read. Therefore, it is the same material that a student is required to read and to be tested on. This was the case with the cloze tests in this study, since both the familiar and unfamiliar materials used in these tests were selected from "Cheng Yu Gu Shi Xuan" which had been provided for CSL learners.

Criterion-related validity coefficients were determined by calculating the correlations between students' cloze test scores and teacher-made classroom test scores as well as students' final grades, because these two tests were the only criteria available for the students' Chinese proficiency.

The Pearson correlation coefficients between the tests are shown in Table 5.

Given the fact that the students' final exam scores and final grades are the criteria, satisfactory validities are noted in cloze tests using familiar materials. That is, the test with every fifth word deletion had a correlation of .83 with the final exam scores and .89 with the final grades. The test with random deletion had a correlation of .87 with the final exam scores and .78 with the final grades.

Students' final exam scores highly correlate with their final grades (r=.89), which indicated that the two criterion tests were consistent and parallel, so that the cloze validity had been calculated on the same standard.

While every fifth deletion test scores correlate better with students' final grades (r=.89, vs. r=.83), random deletion has a higher correlation with the final exam scores (r=.87, vs. r=.78). That is to say, those students who had a higher final grade performed better on every fifth deletion cloze; students who did better on their final exam also had a better score on the random cloze test.

No significant correlation was found between The Nelson-Denny vocabulary test and any of the other tests, which

Table 5
Pearson Correlation Coefficients Matrix
n = 10

	Familiar every 5th deletion	Unfamiliar every 5th deletion	Familiar random deletion	Unfamiliar every 5th deletion	Final exam	Vocabulary test
Unfamiliar every 5th deletion	.66 p=.02	-	-	-	-	-
Familiar random deletion	.77 p=.01	.45 p=.09	-	-	-	-
Unfamiliar random deletion	.28 p=.21	.34 p=.17	.71 p=.01	-	-	-
Final exam	.83 p=.00	.62 p=.03	.87 p=.00	.56 p=.05	-	-
Vocabulary test	-.20 p=.29	-.30 p=.20	-.09 p=.40	.01 p=.50	.21 p=.28	-
Final grades	.89 p=.00	.49 p=.07	.78 p=.00	.36 p=.15	.89 p=.00	.15 p=.34

indicated that the students' Chinese proficiency was not related to their English vocabulary. However, when the vocabulary test score distribution is examined (see Figure 2.), it could be discovered that, despite the large discrepancy between the vocabulary scores (18 points), the majority of the scores clustered around 55 and 60. With this limited band of distribution, a high correlation was hard to obtain.

It is surprising to find that the student with the lowest vocabulary score actually did better on the cloze tests than the student with the highest vocabulary score. The former student's cloze scores on the four cloze tests were: 19, 12, 18 and 14; while those of the latter student were: 7, 14, 13 and 13. Unfortunately, however, a satisfactory explanation could not be found from the information now available.

With unfamiliar materials, tests did not show high correlation with the criterion measure. One explanation of this could be that the subjects had not yet got used to the cloze task and, unfamiliar materials might have caused more confusion.

Compared with the validities reported in the literature (see Table 2), the cloze tests with familiar materials in the present study possess high validities (ranging from .89 to ,78). The validities in those studies ranged from .63 to .89, with the majority below .80.

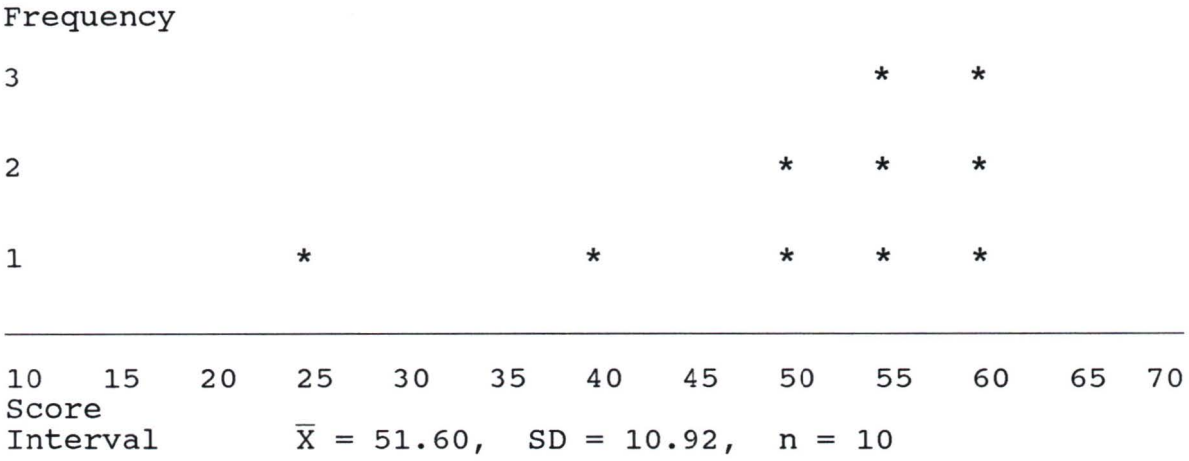

Figure 2
Nelson-Denny Vocabulary Test Score Distribution

Comparison Among the Cloze Tests

The four cloze tests used in this study involved both familiar and unfamiliar materials, and with each type of material, tests were constructed in both every fifth word and random deletion methods. Therefore, it is necessary to investigate the differences between the two types of materials and deletion methods. In order to do this, four pairs of cloze tests were compared and their mean differences are shown in Table 6.

The significant correlation ($r=.77$) indicated a strong relationship between the two tests using familiar materials, although the deletion methods of the two tests were different. Also, there is no significant difference found between the two means ($t=0.36$, $df=9$, not significant). Therefore, these two tests are highly related, which suggests that the every fifth word deletion has the same function as the random deletion method.

When the two tests using the same kind of material were compared, regardless of the different deletion methods they had used, no significant difference was found between the means of the tests using either familiar material ($t=.36$, $df=9$, not significant), or unfamiliar material ($t=.50$, $df=9$, not significant). This suggests that the different deletion

Table 6
Mean Differences of the Four Cloze Tests

Pair of Tests	Mean Score	Std. Dev.	Corre-lation	Correlated t	df	2-tail prob.
Familiar every 5th deletion	15.1	6.64	.66	2.06	9	.07
Unfamiliar every 5th deletion	11.8	3.65				
Familiar every 5th deletion	15.1	6.64	.77	.36	9	.73
Familiar random deletion	14.6	4.12				
Unfamiliar every 5th deletion	11.8	3.65	.34	.50	9	.63
Unfamiliar random deletion	12.4	2.91				
Familiar random deletion	14.6	4.12	.71	2.40	9	.04
Unfamiliar random deletion	12.4	2.91				

methods had not caused any difference to the test results, as far as the mean scores are concerned.

However, different test materials do yield significantly different means. Obviously, tests using familiar materials had higher means than the tests using unfamiliar materials. The difference between the two tests with every fifth deletion method was 3.3 points, and random deletion 2.2 points. The 2.2 points difference is statistically significant ($t=2.40$, $df=9$, $p<.05$). This suggests that the tests using familiar material were at an appropriate level of reading difficulty so that the students performed better in filling in the blanks.

Questionnaire Results

Students' attitudes towards the cloze tests were also important in evaluating the procedure, since it provided some further evidence from a different dimension. Therefore, in order to examine the cloze from the students' perspective, eleven students were asked to respond to an attitude questionnaire following the final administration of the cloze tests. There were five questions to be answered, anonymously (see Appendix B).

The first question was: "What do you think of the cloze test as a measure of your understanding of Chinese?" There were five choices presented as a)very good, b) good, c) neither good nor bad, d) poor, and e) very poor. Six students regarded the cloze test as "good"; three of them chose "neither good nor bad", and only two of them considered the cloze test to be "poor." But none of them responded "very good."

In response to the second question: "In what aspect(s) do you think the cloze test really tests/helps?", "recognizing characters" had been chosen seven times; "reading comprehension", six times; "understanding sentence structure", five times; "understanding word combinations", seven times. No one had chosen "learning grammar" and "learning pinyin."

Students usually don't like any test. If there should be some students who liked to be tested by the cloze test, this could have lent further evidence to the validity or at least utility of the cloze procedure. Therefore, it was quite interesting to ask "Would you like to see the cloze test used regularly as part of an evaluation of your Chinese proficiency?" Seven students gave a negative answer, whereas three of them replied positively. Some students explained that they "do not recognize many characters", or they just "do not have time."

Although it was a subjective question to ask "How much do you think the cloze test reflects your level of Chinese proficiency?", six students considered the cloze test "fairly accurate." Five of them chose "fairly inaccurate."

Only four students wrote extra comments on the use of the cloze test. Comments are listed as follows:

1. "If cloze tests are used they should be typed. Handwriting is an uncontrolled variable."
2. "I've barely, if at all, kept up with the class so I don't think I gave myself a fair chance on these tests."
3. "When key vocabulary in the clozes have not yet been learned, it is sometimes impossible to know what the cloze means and even impossible to guess."
4. "My biggest problem was lack of experience dealing with handwritten Chinese. I had trouble recognizing characters I knew."

It seems that the handwritten tests caused problems to some of the students. Fortunately, by looking at the distribution of the tests using familiar material, all the scores were evenly distributed between the range of 23 and 10, with only one exceptional score on 2. No exceptional score was found in unfamiliar material test scores, which ranged from 4 to 18. In addition, all the mean scores of the cloze tests were above 44% frustration level(p.42), which, according to Bormuth (1968), is equal to a 75% comprehension score. Therefore, handwritten characters had not really caused fundamental effects on the cloze test results as a

whole, although a few students claimed to have difficuty in recognizing them. Nevertheless, in order to ensure the reliability and validity of the test, a typed or printed format should be used in a formal testing procedure.

On the whole, the results of the questionnaire showed that most of the students had a positive view towards the cloze tests, although some of them didn't want it to be used as a regular evaluation instrument, which, in the author's opinion, might be a purely psychological refusal towards tests in general, rather than a result of the poor quality of this particular test. It also shows that the students had a fairly correct understanding of the test, except some of them still thought that the cloze test could be done only after every character in the passage had been learned. However, the nature of the cloze procedure requires the subjects to consider the cloze pattern as a whole and make intelligent guesses.

Chapter 6

CONCLUSIONS

Summary

This study was designed to determine the effectiveness of the cloze test as a measure of Chinese language proficiency.

Regarding the research questions, the study found that cloze tests using familiar materials had a high correlation with both the students' final exam scores and their final grades, which were considered the criteria of the students' Chinese proficiency. Hence, it was then believed that these tests possess satisfactory validity. And cloze tests using familiar materials also yielded high reliability. Therefore, only the cloze tests using familiar materials appear to validly and reliably test the CSL proficiency of the students at this particular level.

There was no significant difference between the mean scores of the tests using random deletion and the every fifth word deletion method. Thus, it confirmed the close relationship between the two deletion methods.

No significant correlation was found between the English vocabulary test scores and the students' Chinese proficiency. Hence, the cloze test using familiar material is a better predictor of the students' Chinese proficiency than the English vocabulary reading level test.

The questionnaire also revealed that the students involved in this study had a clear understanding and a positive attitude towards the cloze test. They also indicated an unexpected problem - handwritten Chinese, which had been an uncontrolled variable in the study.

Limitations of the Study

1. A major limitation of this study was the availability of subjects. It was not possible to acquire a large number of subjects in Victoria who were learning Chinese as a second language at a third or fourth year level. This resulted in a sample of only ten subjects. This might have reduced the degree to which the results can be generalized.

2. Although the unfamiliar cloze passages were selected in accordance with the textbook and the students' general reading ability, differences would still exist between the familiar and unfamiliar cloze passages in terms of reading

difficulty. Therefore, this unequal difficulty might have affected the students' performance.

3. The availability of a standardized Chinese proficiency test would have provided additional evidence of the validity of the tests used in this study.

4. The handwritten test format could have influenced students' performances.

5. The addition of the Nelson-Denny Comprehension sub-test might have provided a better measure of students' reading level than the Vocabulary sub-test alone.

Implications for Future Research

The present study explored the possibility of using cloze procedure as a measure of students' proficiency of Chinese as a second language. Although the cloze has been used to test English proficiency for nearly 40 years, little has been done with tests in Chinese. It is certainly of significance to explore its function further in novel situations.

Due to the limitations of the study, it could only deal with a few issues in this area. Therefore, the following recommendations are made for future study.

1. A study using a larger sample should be introduced for this kind of research. Subjects should be randomly assigned into several groups, with each group tested on a single cloze format. Try to use the same raw material to construct different cloze formats among the groups, so that the reading difficulty of the passages could be controlled.

2. A longer passage is preferred so that the suggested fifty blanks within a single test could be created in order to further improve the reliability of the test.

3. The test format must be printed in standard Chinese characters, as would be found in the students' textbooks.

4. It might be useful to conduct more research in this field. The cloze test could also be used as a formative evaluation, such as a teaching technique, as well as a sumative testing instrument. The effectiveness of the cloze should be evaluated through long term observation.

5. The present study failed to find out at what level unfamiliar materials could be used. Future research should be conducted at several age groups to find an answer to this problem.

6. This study dealt with the cloze test in relation to the Chinese language. It is then necessary to conduct extended research into the Chinese language itself in terms

reading psychology, with an emphasis on the nature of logographic scripts vs. alphabetic scripts. The ultimate purpose of this is to establish its own theoretical basis for Chinese cloze test and eventually a proper modification of the English cloze format could be found for the Chinese cloze test. It is then expected that Chinese cloze test can be established as a sound and sophisticated testing instrument.

Concluding Statement

This study developed a set of cloze tests to determine the students' proficiency in learning Chinese as a second language. With familiar materials, cloze procedure was found to be a reliable and valid measure of the students' Chinese proficiency. Due to the limitations of the present study, only a few aspects have been dealt with in this virgin soil. Nevertheless, however, the primary purpose of the study will be achieved if it could be recognized as a successful beginning of a pioneer investigation.

REFERENCES

Aitken, K.G. (1977). Using cloze procedure as an overall language proficiency test. TESOL Quarterly 2(1), 59-67.

Alderson, J. Charles. (1978). A study of the cloze procedure with native and non-native speakers of English. Unpublished Ph.D. Dissertation, University of Edinburgh, Scotland.

Alderson, J. Charles. (1979a). Scoring procedures for use on cloze tests. Paper presented at the 1979 TESOL Convention, Boston, Mass.

Alderson, J. Charles. (1979b). The cloze procedure and proficiency in English as a foreign language. TESOL Quarterly, 13, 219-227.

Alderson, J. Charles. (1979c). The effect on the cloze test of changes in deletion frequency. Journal of research in reading, 2, 108-119.

Anderson, J. (1971). Selecting a suitable "reader": Procedures for teachers to assess language difficulty, RELC Journal, 2, 35-42.

Anderson, J. (1976). Psycholinguistic experiments in foreign language testing. St. Lucia, Queensland: University of Queensland Press.

Anderson, J. (1983) The cloze procedure and proficiency in English as a foreign language. In J. W. Oller, (Ed.), Issues in language testing research. (pp.205-217) Rowley, MA: Newbury House.

Borg, R. W., & Gall, D. M. (1989). Educational research: An introduction, (5th Ed.), New York: Longman Inc.

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

Bormuth, J. R. (1969). Empirical determination of the instructional reading level. Reading and realism, 13(1), 716-721.

Briere, E. J., Clausing, C., & Pincell, E. (1978). A look at cloze testing across language and levels. Modern Language Journal, 62, 23-26.

Brown. J. D. (1980). Relative merits of four methods for scoring cloze tests, The Modern Language Journal, 64(1), 311-317.

Brown. J. I. (1973). The Nelson-Denny Reading Test. Boston: Houghton Mifflin Company.

Chance, L. L. (1988). Cloze tests and teacher-made tests: A comparison. Ohio Reading Teacher, 23(1), 6-10.

Chavez-Oller, A. M., Weaver, T., & Oller, John W. (1985). When are cloze items sensitive to constraints across sentences? Language Learning, 35, 181-206.

Conrad, C. A. (1970). The cloze procedure as a measure of English proficiency. MA Thesis. University of California, Los Angelas.

Culhane, J. (1970). Cloze procedure and comprehension. Reading Teacher, 23(5), 410-413.

Darnell, D. K. (1968). The development of an English language proficiency test of foreign students, using a clozentropy procedure. (ERIC Document Reproduction Service No. ED 024 039)

Darnell, D. K. (1970). Clozentropy: A procedure for testing the English language proficiency of foreign students. Speech Monographs, 38, 350-353.

DeSanti, R. J. (1989). Concurrent and predictive validity of a semantically and syntactically sensitive cloze system. Reading Research and Instruction, 28, 29-40.

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

Gronlund, W. E. (1990). Measurement and evaluation in teaching (6th ed.). New York: Macmillan Publishing Company.

Guthrie, E. R. (1952). The psychology of learning. (Rev.ed.). New York: Harper & Row.

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

Hartmann, G. W. (1935). Gestalt Psychology: A survey of facts and principles. New York: The Ronald Press Company.

Helfeldt, J. P., & Henk, W. A. (1985). Usefulness of conventional vs. total random cued cloze tests as measures of reading comprehension. Journal of Reading, 28(8), 719-725.

Henk, W. A., & Helfeldt, J. P. (1985). Effects of alternative deletion patterns, blank conditions, and scoring systems in intermediate level cloze test performance. Reading Psychology, 6(1-2), 85-96.

Hill, W. F. (1971). Learning: A survey of psychological interpretations. New York: Chandler Publication Company.

Hinofotis, F. (1976). An Investigation of the concurrent validity of cloze testing as a measure of overall proficiency in English as a second language. Doctoral Dissertation, Southern Illinois University

Hinofotis, F., & Snow, B. (1980). An alternative cloze testing procedure: Multiple-choice format. In John W. Oller & Kyle Perkins (Eds.), Research in Language Testing, pp.129-33, Rowley, MA: Newbury House.

Irvine, P., Atai, P. & Oller, J. W. Jr. (1974). Cloze, dictation and the test of English as a foreign language. Language Learning, 24, 245-52.

Jongsma E. (1971). The cloze procedure: A survey of the research. (ERIC Document Reproduction Service No. ED 050 893)

Jonz, J. (1975). Values of hierarchies of three groups of Spanish-English multilingual adolescents. Doctoral Dissertation, University of New Mexico.

Jonz, J. (1976). Improving on the basic egg: The M C cloze. Language Learning, 26, 255-265.

Kamil, M. L., Smith-Burke, M., & Rodriguez-Brown, F. (1986). The sensitivity of cloze to intersentential integration of information in Spanish bilingual populations. In Jerome A. N. & Rosary V. L. (Eds.) Solving Problems in Literacy: Learners, Teachers, and Researchers. pp.334-338. Thirty-fifth Yearbook of The National Reading Conference.

Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. Psychometrika, 2, 151-160.

Liu, Y. L. (1989). Issues on testing Chinese as a second language. In Y. L. Liu (Ed.), Research in Chinese Proficiency Test. (pp.1-31) Beijing: Modern Publishing Company.

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

MacGinitie, W. (1961). Contextual constraint in English prose paragraphs. Journal of Psychology, 51, 121-130.

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

McKenna, M. C., & Robinson, R. D. (1980). An Introduction to the Cloze Procedure. International Reading Assosiation, Newark, Delaware.

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

Ohnmacht, F. W., Weaver, W. W., & Kohler, G. T. (1970). Cloze and closure: A factorial study. The Journal of Psychology 74, 205-17.

Oller, J. W. (1972). Scoring methods, and difficulty levels for cloze tests of ESL proficiency, Modern Language Journal, 56, 151-158.

Oller, J. W. (1979). Language tests at school. London: Longman Group Limited.

Oller, J. W., & Conrad, C. A. (1971). The cloze technique and ESL proficiency. Language Learning, 21(2), 183-195.

Pike, L. W. (1973). An evaluation of present and alternative item formats for use the test of English as a foreign language. Princeton, N.J.: Educational Testing Service.

Rankin E. F. Jr. (1970). The Cloze Procedure---Its Validity and Utility. In Roger Farr, Harcourt (Ed.), Measurement and Evaluation of Reading. New York: Brace and World, Inc.

Rye, J. (1982). Cloze procedure and the teaching of reading. London: Heinemann Educational Books Ltd.

Soudek, M & Soudek, L. I. (1983). Cloze after thirty years: New uses in language teaching. ELT Journal, 37, 335-9.

Spolsky, B. (1969). Recent research in TESOL. TESOL Quarterly, 3, 355.

Spolsky, B. (1971). Reduced redundancy as a language testing tool. In G. E. Perre and J. L. M. Trim (Eds.). Application of linguistics, Cambridge: Cambridge University Press, pp.385-390.

Stephens, E. C., & Weaver, D. R. (1986). The cloze procedure as predictor of undergraduate achievement in introductory courses. Forum for Reading, 18(2), 32-36.

Stubbs, J. B., & Tucker, G. R. (1974). The cloze test as a measure of English proficiency, Modern Language Journal, 58, 239-241.

Syme, R. M. (1989). Random cloze and an alternative to the traditional cloze. MA Thesis, University of Victoria.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability, Journalism Quarterly, 30, 414-438.

Taylor, W. L. (1954). Application of "cloze" and entropy measurs to the study of contextual constraint in samples of continuous prose. Doctoral dissertation, University of Illinois.

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

Travers, R. M. W. (1982). Essentials of learning. New York: Macmillan Publishing Co., Inc.

Van Every, K. (1981). Use of the cloze to predict ease of success in a mastery learning situation, Teaching of psychology, 8, 114-115.

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

Wijnstra, J. M., & Wageningen, N. (1974). The cloze procedure as a measure of first and second language language proficiency. Mimeograph.

Appendix A - Cloze Tests Used in the Study

Practice Test

Directions: Read the whole passage through. Then, go back and fill in the blanks with characters you think are missing. Use only one character for each blank. Credit will also be given for pinyin.

从前，有个___肚子饿了，就卖了五___包子吃起来。一个、___个、三个、四___……。他吃___四个包___以后，肚子还是___。

他想："怎___办？只剩___一个包子___。"他把最后___一个包子___了以后，肚___才饱了。

他又想："唉！我___笨，要是早___道第五个___子可以___得饱，我刚才为___么要吃前___的那四___呢？"

Cloze Test 1
(every 5th word deletion using familiar material)

Directions: Read the whole passage through. Then, go back and fill in the blanks with characters you think are missing. Use only one character for each blank. Credit will also be given for pinyin.

东施见西___长得漂亮，___想跟西施___。西施怎么___扮，她也就___么打扮；西___怎么走路，___也就怎么___路。西施有___痛病，发起___来，总是紧___皱起眉头，___手按住胸___。

东施为了___别人说她___看，也学着___施的样子，___紧皱起眉___，用手按住___口，以为这___就同西施___样漂亮了。___是没有人___她好看，反___说她越学___难看。

Cloze Test 2
(every 5th word deletion using unfamiliar material)

Directions: Read the whole passage through. Then, go back and fill in the blanks with characters you think are missing. Use only one character for each blank. Credit will also be given for pinyin.

古时候，有___个朋友在___起喝酒。他___人很多。可___酒太少，只___一壶酒。这___酒给谁喝___？有人说："我___每人都画___条蛇，大家___一比，看谁___得快。画得___快的人喝___壶酒，好吗？

___家都说："好"。

___们开始画___了。有一个___轻人比别___画得快，他___别人还在___呢，就高兴___说："你们画___太慢了！我___有时间，让___再给蛇加___只脚吧。"

Cloze Test 3
(random deletion using familiar material)

Directions: Read the whole passage through. Then, go back and fill in the blanks with characters you think are missing. Use only one character for each blank. Credit will also be given for pinyin.

从前，楚国有___个卖兵器的___，在市场上出卖___和盾。他举起___，向人___夸口说："___的盾，坚固无___，世上任___锋利的东___都不能刺穿___"。

卖兵___的人又挥舞着___的长矛，向人___吹嘘着："___这枝矛锋利___比，无___怎样坚硬的___，一碰上，就能戳___！"

人群中有___问道："___拿你的矛来刺___的盾吧，看___结果怎___样。"

那个卖兵器的___张口结舌，无从回答，___着矛和盾走了。

Cloze Test 4
(random deletion using unfamiliar material)

Directions: Read the whole passage through. Then, go back and fill in the blanks with characters you think are missing. Use only one character for each blank. Credit will also be given for pinyin.

从前，有一个___，养了十几只羊。一___早上，他去放羊，发现___少了一只。原来___圈破了___个洞。晚上狼从这___洞进去，吃了___只羊。

邻居劝___说："赶快___羊圈修___修吧！"

他___："羊已经丢了，修理___圈还有什么___呢！"

第二天早___，他发现又少了一___羊，狼又___这个洞进到羊___里，吃___一只羊。

这时，他明___了。邻居___意见___对的，他赶快___羊圈修好了。以___，他的羊再也没___被狼狐走___。

Appendix B

QUESTIONNAIRE

1. What do you think of the cloze test as a measure of your understanding of Chinese?
   a) very good b) good c) neither good nor bad
   d) poor e) very poor

2. In what aspect(s) do you think the cloze test really tests/helps? (You can choose more than one of the following.)
   a) recognizing characters b) reading comprehension
   c) learning grammar d) learning pinyin
   e) understanding sentence structures
   f) understanding word combinations

3. Would you like to see cloze test used regularly as part of an evaluation of your Chinese proficiency?
   a) Yes b) No
   If no, why?

4. How much do you think the cloze test reflects your level of Chinese proficiency?
   a) Very accurate b) Fairly accurate
   c) Fairly inaccurate d) Very inaccurate

Have any other comments on the use of cloze tests in measuring your proficiency with Chinese?

VITA

Surname: Xie Given Names Wang

Place of Birth Shanghai, People's Republic of China

Date of Birth January 29, 1964

Educational Institutions Attended:

East China Normal University	1981 to 1985
University of Victory	1989 to 1991

Degrees Awarded:

B.A. (English language and Literature)	1985

Honour and Awards:

IDRC Fellowship	1989 to 1991

Publications:

Xie, W. (1987). An Abridged Report on a Sample Survey of the 1986 Shanghai College Entrance Examination in English, English Teaching and Research Notes (1), 35-37

Xie, W. (1989). English Standardized Test: Its Basic Concepts, Preparation and Implications, English Teaching and Research Notes (3), 14-17, 36.

PARTIAL COPYRIGHT LICENSE

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

Title of Thesis: Cloze Procedure as a Proficiency Test of Chinese as a Second Language

Author 

WANG XIE

July 20, 1991