

AN EXPERIMENTAL STUDY OF THE  
MULTIPLE - CHOICE TEST  
TECHNIQUE FOR MEASURING  
READING COMPREHENSION OF EFL STUDENTS

A THESIS

Submitted to the Institute of Economics and Social  
Sciences of Bilkent University  
in Partial Fulfillment of the Requirements  
for the Degree of Master of Arts  
in the Teaching of English as a Foreign Language

By

DEMET ÇELEBİ  
AUGUST 1991

## ABSTRACT

Since reading is an integrative and difficult skill to measure no single test adequately measures reading ability. Testing reading comprehension is problematic, especially in a foreign or second language setting, because test results may be invalid; a reading comprehension test score may be measuring something other than the reading comprehension ability of the student. The purpose of this study is to test two reading comprehension testing techniques, multiple-choice and retelling by conducting an experiment using intermediate level EFL students from Bilkent University School of English Language (BUSEL) as subjects. The goal of the experiment is to correlate these two testing techniques with two other measures of reading comprehension ability, teacher ratings and the reading section of a BUSEL achievement exam. The correlations of these four measures were done using the Pearson Product Moment Correlation and the Tetrachoric Correlation. The results showed that the multiple-choice reading test did not correlate significantly with the other three measures, implying that the multiple-choice test did not successfully predict the reading abilities of students. On the other hand, the correlation of the retelling procedure with the other measures showed significance,

suggesting that the retelling is a more valid way to test reading comprehension. Because the total number of subjects in this study was small, only 18, we should hesitate before generalizing and applying these results to other EFL situations. However, the findings of the experiment should be of interest to all those involved in testing the reading comprehension ability of foreign or second language students.

AN EXPERIMENTAL STUDY OF THE MULTIPLE-CHOICE TEST  
TECHNIQUE FOR MEASURING  
READING COMPREHENSION OF EFL STUDENTS

A THESIS  
SUBMITTED TO THE INSTITUTE OF ECONOMICS AND SOCIAL SCIENCES  
OF BILKENT UNIVERSITY  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF MASTER OF ARTS  
IN THE TEACHING OF ENGLISH AS A FOREIGN LANGUAGE

BY

DEMET CELEBI

AUGUST 1991

*Demet Celebi*  
1991

PE  
1066  
.C45  
1991  
B.1268

BILKENT UNIVERSITY  
INSTITUTE OF ECONOMICS AND SOCIAL SCIENCES  
MA THESIS EXAMINATION RESULT FORM

July 31, 1991

The examining committee appointed by the  
Institute of Economics and Social Sciences for the  
thesis examination of the MA TEFL student

Demet Celebi

---

has read the thesis of the student.  
The committee has decided that the thesis  
of the student is satisfactory.

Thesis title : An Experimental Study of the  
Multiple-Choice Test  
Technique for Measuring Reading  
Comprehension of EFL Students.

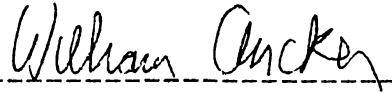
Thesis Advisor : Mr. William Ancker  
Bilkent University, MA TEFL Program

Committee Members : Dr. James C. Stalker  
Bilkent University, MA TEFL Program

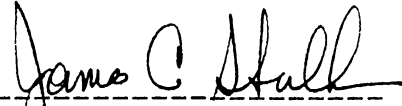
Dr. Lionel Kaufman  
Bilkent University, MA TEFL Program



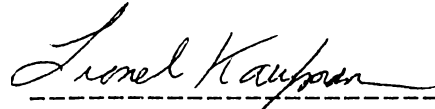
We certify that we have read this thesis and that in our combined opinion it is fully adequate, in scope and in quality as a thesis for the degree of Master of Arts.



William Ancker  
(Advisor)

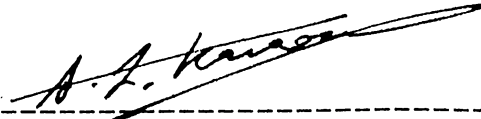


James C. Stalker  
(Committee Member)



Lionel Kaufman  
(Committee Member)

Approved for the  
Institute of Economics and Social Sciences



Ali Karaosmanoglu  
Director  
Institute of Economics and Social Sciences



## TABLE OF CONTENTS

SECTIONS	PAGES
1.0 INTRODUCTION	
1.1 Background of the Study	1
1.2 Statement of the Research Question	2
1.3 Hypotheses	3
1.3.1 Null Hypothesis	3
1.3.2 Experimental Hypothesis	3
1.4 Purpose and Methodology of the Study	3
1.5 Organization of the Study	5
2.0 REVIEW OF LITERATURE	
2.1 Reading Comprehension	6
2.2 Testing Reading Comprehension	9
2.3 Multiple-Choice Tests	10
2.4 Some Weaknessess and Strengths of Multiple-Choice Tests	12
2.5 Suggestions for Preparing Better Multiple-Choice Reading Tests	15
2.5.1 Selection of Multiple-Choice Distractors	15
2.5.2 How to Improve Multiple-Choice Reading Tests	16

2.5.3	An Alternative Approach to Multiple-Choice Reading Items	17
2.6	Retelling - A More Direct Way of Testing Reading	18
2.7	Conclusions	19
3.0	METHODOLOGY	
3.1	Introduction	20
3.2	Participants and Setting	21
3.2.1	Subjects	21
3.2.2	Experimenters	22
3.3	Materials	23
3.4	Two Outside Measures Used as the Criteria for Measuring Reading Ability	24
3.5	Procedures of Data Collection	26
4.0	DATA ANALYSIS	
4.1	Introduction	28
4.2	Data Analysis	29
4.2.1	Pearson Product Moment Correlation	31
4.2.2	Tetrachoric Correlation	33
4.3	Results	40
5.0	CONCLUSIONS	
5.1	Summary of the Study	42

5.2	Assessment of the Study	44
5.3	Pedagogical Implications	46
5.4	Implications for Future Research	47
	BIBLIOGRAPHY	50
	APPENDIX A	52
	APPENDIX B	53
	APPENDIX C	54
	APPENDIX D	56
	APPENDIX E	58
	APPENDIX F	59
	APPENDIX G.1,2,3,4,5,6	60
	APPENDIX H.1,2,3	66

## LIST OF TABLES

TABLES	PAGES
Table 4.1 PPMC Calculations Between the Four Tests	31
Table 4.2 Tetrachoric Correlation Between the Multiple-Choice Test and the Teachers' Rating	37
Table 4.3 Tetrachoric Correlation Between Retelling and the Teachers' Rating	38
Table 4.4 Tetrachoric Correlation Between the BUSEL Reading Exam and the Teachers' Rating	40

## ACKNOWLEDGEMENTS

I am indebted to my advisor Mr. William Ancker for his guidance, feedback and encouragement while writing this thesis.

I would like to express my thanks to Dr. James C. Stalker and Dr. Lionel Kaufman for their valuable comments and professional assistance.

I owe special thanks to my colleagues at BUSEL and my classmates in the MA TEFL Program. Without their help it would have been impossible to carry out this study.

My sincere thanks go to my classmate Aysel Sutcu for her endless moral support and cooperation.

I would also like to express my deep appreciation to my father and my brother-in-law for their assistance in using the computer.

CHAPTER I  
INTRODUCTION

1.1 Background of the Study

In English-medium universities reading comprehension is considered to be one of the most important skills as students are required to do all their academic reading in English. Reading comprehension tests are used in assessing the comprehension of readers in general and EFL students. There are a variety of tests that are used in measuring different reading skills for example, pictures and sentence matching, true-false tests, cloze tests, multiple-choice tests, and retelling. Among these tests multiple-choice tests are widely used as objective tests, especially to measure reading comprehension, but whether or not these tests assess comprehension adequately is a question. According to Aslanian (1985), multiple-choice tests are an inadequate means of assessing the comprehension of readers in general and of ESL/EFL students in particular. Aslanian states that comprehension is a complex phenomenon and cannot be measured by objective tests alone; discussing the weaknesses of multiple-choice reading tests, she says:

Among the drawbacks inherent in multiple-choice reading tests is the fact that most provide sample passages that are too truncated or incomplete to resemble the normal printed

matter available outside the classroom. These passages often lack a proper introduction, sufficient content development, and a coherent conclusion. To expect a reader, especially an ESL reader, to understand the omissions that have been made to render the test economical is to expect confusion, not comprehension. (1985, p. 1)

Heaton, on the other hand, argues that a " multiple-choice test offers a useful way of testing reading comprehension, and the extent to which a test is successful in measuring what it sets out to measure depends largely on the effectiveness of each of the items used" (1988, p. 117).

Unlike multiple-choice, which is a recognition type of test and the results of which do not necessarily show comprehension, the retelling procedure is a more direct and valid way of testing reading comprehension as here the testee has to repeat the text in his/her own words. In the retelling procedure the testee will either be able to interpret the text correctly or not, which will show whether comprehension took place.

## 1.2 Statement of Research Question

The purpose of this study was to find out how well multiple-choice testing and retelling testing measure students' reading comprehension. To find out the validity of these two reading tests they were compared and correlated with teacher ratings and BUSEL reading

exam results.

### 1.3 Hypotheses

#### 1.3.1 Null Hypothesis

a. There is no correlation between the multiple-choice reading test and the teacher ratings and the BUSEL reading exam.

b. Similarly, there is no correlation between the retelling procedure and the teacher ratings and the BUSEL reading exam.

#### 1.3.2 Experimental Hypothesis

a. There is a positive correlation between the multiple-choice reading test and the teacher ratings and the BUSEL reading exam.

b. There is a positive correlation between the retelling procedure and the teacher ratings and the BUSEL reading exam.

### 1.4 Purpose and Methodology of the Study

Multiple-choice tests are widely used in the preparatory schools of English-medium universities, including Bilkent University School of English Language (BUSEL), for assessing reading comprehension without really questioning the validity of these tests as measures of comprehension. This study aims at finding how well the multiple-choice reading test and the



retelling procedure measure students' comprehension of a reading passage. In order to find out whether the multiple-choice reading test or the retelling procedure (or both tests) are appropriate means of assessing reading comprehension, they were compared and correlated with the teacher ratings (see Chapter IV, section 4.2.3) and the results of the BUSEL reading exam (see Appendix A). So testing the comprehension of a reading text through a multiple-choice test and testing the same text through retelling were correlated with the two outside measures, teacher ratings and BUSEL reading exam, to find out which of the two reading testing techniques best measure the reading abilities of students.

The study was conducted with 18 intermediate level Bilkent University School of English Language students. The students' two class teachers were asked to rate each of the students' reading abilities on a reading rating scale. In addition, the most recent BUSEL reading exam results were used. These two ratings were taken as the criteria for determining whether the two tests - multiple-choice and retelling - measure students' reading abilities. Subjects were first given a reading passage with 8 multiple-choice items and were asked to complete it (see Appendix B). After the first task was completed subjects were asked to retell the reading passage by repeating their understanding of the passage in their own

words. Subjects were then graded on these two different types of tests and the scores obtained were compared. A correlational study was done between these two tests and the teacher ratings and the BUSEL reading exam as to which test was a better means of assessing reading comprehension.

### 1.5 Organization of the Study

Chapter II - This chapter gives a review of literature about reading comprehension, multiple-choice tests, and retelling.

Chapter III - This chapter gives a detailed explanation of the methodology used for collecting and analysing this data.

Chapter IV - In this chapter data is presented and the procedure carried out to analyse data is explained step by step. This chapter also gives an idea about the results of the study.

Chapter V - This final chapter gives an overall summary of the study focusing on the results. In this chapter the hypothesis of the study and the obtained result and the implications for further study are discussed.

CHAPTER II  
REVIEW OF LITERATURE

2.1 Reading Comprehension

Reading is one of the most important skills that is taught at English-medium universities. Teaching this skill is given great emphasis but first of all it is important to understand what reading is in order to teach it successfully. There are several definitions of reading. According to Riley (1977) reading is a receptive skill, in the sense that, the reader is expected to decode the visual form of the text into comprehensible messages in the brain. The reader has to integrate his reading abilities in order to decode the message correctly. Riley (1977, p. 3) quoting from Goodman further explains the concept as:

The reader associates the semantic properties of language to form which appears on the paper as an integration of all grammatical units of a language, since the written form and the spoken form of English is rather different, a phonemic correspondence is well necessary to comprehend the written form.

As pointed out above, to decode the message and to read for comprehension, the reader has to use a number of subskills during the reading process. According to Riley (1977, p. 4), an intelligent reader has to be able to use the following subskills while reading for comprehension:

- 1) Select main idea from the passage
- 2) Select relevant details to support the main

idea

- 3) Recognize irrelevancies, contradictions and non sequiturs
- 4) Use logical connectors and sequence signals
- 5) Draw conclusions
- 6) Make generalizations
- 7) Apply principles to other instances.

All these subskills constitute a whole in comprehending a written text and the testee is exposed to a similar process in a reading comprehension test.

Goodman (qtd.in Singer and Ruddell, 1985) on the other hand, defines comprehension as the end product of any act of reading, and shows how comprehension may change in the course of testing for it:

Since comprehension is a constructive process in which readers make sense of the text, it goes on during reading and even long afterwards as the reader reconsiders and reconstructs what has been comprehended; thus, comprehension may be changed in the course of testing it. The reader may change what he or she understood on the basis of the test questions which seem to require particular responses and views. (p. 831)

Galloway (1988) looks at comprehension in terms of how well the purpose of the reader in approaching the text corresponds with the purpose of the writer in preparing it. Valette (1977) has a similar point of view and thinks that the main objective of reading for comprehension is to understand the written message. If reading is considered as a communicative activity, then, the realization of communication in reading is the moment when the reader understands the message the writer wants

to convey.

Johnston (1983) examines reading from another perspective and sees it as, "any reader interaction with the text" (p. 2). He states that a reader comprehends a text after interacting with it. Comprehension is stored in the reader's memory and can be examined by getting the reader to express segments of the stored material. He further points out that the processes involved in getting there are given less emphasis than the final product (i.e. the contents of memory). The product implies that long-term memory has an important role in comprehension and determines how successful the reader is at comprehending. According to Johnston, (1983) comprehension can be measured by standardized tests and free recall measures.

In contrast, Carroll (1971, ctd. in Johnston, 1983, p. 6) argues that comprehension is a process that takes place on reception of information and that only short term memory is involved. He states, "As soon as longer intervals are involved in the testing of comprehension, there is the possibility that we are studying memory processes along with, or in place of, comprehension processes".

Royer and Cunningham (1978, in Johnston, 1983) take issue with both the process and product approaches and contend that "... comprehension processes and memory

processes are inextricably intertwined...We assume that a comprehended message will be retained in memory better than an uncomprehended message" (p. 2).

Pearson and Johnson (1978, qtd. in Johnston, 1983) believe that:

Comprehension is building bridges between the new and the known...comprehension is active not passive; that is, the reader cannot help but interpret and alter what he reads in accordance with prior knowledge about the topic under discussion. Comprehension is not simply a matter of recording and reporting verbatim what has been read, comprehension involves a great deal of inference making. (p. 24)

## 2.2 Testing Reading Comprehension

Because reading and comprehending what is read are internal processes and takes place in the brain, they are difficult to observe and measure. Research does not allow us to say that there is a single testing technique that is the most appropriate and reliable way of testing reading comprehension. Madsen (1983) agrees with the idea that reading comprehension tests are difficult and says: " Testing reading comprehension is one of the most integrative and challenging kinds of reading test types" (p. 87).

Oller (1979) also contributes to the idea that reading tests are integrative and challenging and says that these tests require the testee to read and answer questions - but they usually leave open the question of

what the test is a test of. A reading test may be a good test of overall language proficiency or a test of listening comprehension rather than being a good reading comprehension test.

There are many ways to test reading comprehension. One of the most common and objective ways is a reading passage followed by multiple-choice questions. As mentioned above, comprehension is viewed as a bridge between the new and the known, but Wood (1987) does not think that this is the case especially with multiple-choice reading tests. She states her view as follows:

In doing a multiple choice test however, the reader does not necessarily build a bridge between his past experience and background information and the new. Readers must see what the tester wants them to see, which is often literal factual information, easy to locate in the text without careful reading. Therefore, these texts are often criticized for testing little more than the reader's ability to answer multiple choice questions.  
(p. 4)

### 2.3 Multiple-Choice Tests

Multiple-choice items take many forms, but the initial part of each multiple-choice item is known as the stem; the choices from which the students select their answers are referred to as options/responses/alternatives. One option is the answer/correct option/key, while the other options are distractors. The task

of a distractor is to distract the majority of poor students (i.e those who do not know the answer) from the correct option (Heaton, 1988, p. 28).

Heaton (1988) suggests some guidelines on preparing multiple-choice tests of reading. The reading passage used in the test might vary from 50 to 100 words at the elementary level, 200 to 300 words at the intermediate level, and 400 to 600 words at the advanced level. He goes on to say that the passage must be able to provide a sufficient number of multiple-choice comprehension items and the length of the passage should be related to its level of difficulty. Deciding on the number of items depends on the length and complexity of the text. Testees need much more time to complete a reading comprehension multiple-choice test than a multiple-choice vocabulary test as they first have to read the passage and then answer the questions.

The number of alternatives, or options, for each multiple-choice item is five in most public tests but as it is extremely difficult to construct four equally good distractors, four options are recommended for most classroom tests, three being the distractors and one being the correct answer. If a multiple-choice reading comprehension test has been carefully prepared, then, the choice of the correct option must depend on a testee's comprehension of the reading text rather than on general



knowledge or intelligence.

Extensive research has been done in this field of testing reading, revealing both weaknesses and strengths in the use of multiple-choice tests.

#### 2.4 Some Weaknesses and Strengths of Multiple-Choice Tests

The construction of multiple-choice tests is generally complicated and difficult. Bensoussan (1984) argues that multiple-choice tests are difficult to construct and that they do not require the student to write in order to provide evidence of his reading comprehension. Oller (1979) does not favor multiple-choice tests and sees a lot of problems in constructing and administering these tests. He lists a number of principles that should be taken into consideration while preparing multiple-choice tests. For example, are the items in the test related to the skill, construct, or curriculum that the test is supposed to assess? Are the items appropriately chosen and the distractors attractive enough to distinguish between the good and the poor learners? Is the test asking the testees to guess facts that are not stated or implied? Are there clues about the correct answer in the options given? The only values that Oller (1979) sees in using multiple-choice tests are that they are objective and that they are easy and

reliable to score. Also when large numbers of people are to be tested, multiple-choice tests are very economical. Oller believes in the practicality of multiple-choice tests, but does not believe in their reliability or validity and does not recommend the instructional use of multiple-choice tests.

Hughes (1989, p. 59) shares the same opinion as Oller and says that, "the most obvious advantage of multiple-choice is that scoring can be perfectly reliable, rapid and economical". He also sees limitations and difficulties with multiple-choice tests:

If there is a lack of fit between at least some candidates' productive and receptive skills, then performance on a multiple-choice test may give a quite inaccurate picture of those candidates' ability ... The chance of guessing the correct answer in a three-option multiple-choice item is one in three, or roughly thirty-three per cent. On average we would expect someone to score 33 on a 100-item test purely by guesswork. We would expect some people to score fewer than that by guessing, others to score more. The trouble is that we can never know what part of any particular individual's score has come through guessing. (p. 61)

In multiple-choice tests, item writing is one of the most difficult aspects. Hughes (1989) does not recommend multiple-choice tests for regular achievement testing within institutions because he thinks that the amount of work and expertise needed to prepare good multiple-choice tests are so great that it is not worth it.

Madsen (1983) lists three objections to multiple-

choice tests. First, he believes that multiple-choice tests are passive and rely on recognition rather than production. Secondly, they expose students to errors, and finally they may indicate that a student who has just guessed on the test is successful.

Aslanian (1985) also does not favour multiple-choice tests and believes that these tests are an inadequate means of assessing the comprehension of readers in general and ESL students in particular. She states that comprehension is a complex phenomenon and cannot be measured by objective tests alone. Aslanian argues that multiple-choice tests often measure what the test-makers believe is important, not what the student understands or thinks is important in the reading passage. She also believes that in multiple-choice comprehension tests the correct response is often not selected by searching for the meaning in the passage but by eliminating the obviously unacceptable or absurd distractors.

Although there is controversy about multiple-choice reading tests, many teachers like multiple-choice reading tests and think that they reflect the receptive nature of reading. These teachers do not believe that it is possible to pass an objective test just by guessing.

## 2.5 Suggestions for Preparing Better Multiple-Choice Reading Tests

Multiple-choice tests are objective, easy to score, therefore economical. These tests are widely used and seem to be here to stay. Instead of merely accepting the shortcomings of these tests, however, we should try to deal with them and find solutions to make them more valid, that is, make them really measure what they are said to measure.

### 2.5.1 Selection of Multiple-Choice Distractors

As mentioned before, the most difficult area in constructing multiple-choice tests is the selection of distractors. It is often the teachers who choose the distractors and they usually use their intuition while preparing them. Cohen (1980) thinks that problems might arise in that a teacher's intuitions about what will "distract" students away from the correct answer might not always be accurate.

Dobson (1974, in Cohen, 1980) researched the possibility of obtaining distractors for multiple-choice items from among high frequency incorrect answers on completion items given to students. She showed that distractors prepared from incorrect student answers for distinguishing good from bad readers were more reliable. In her study, a multiple-choice test with distractors obtained from analysis of answers to completion exercises

was found to be more reliable, and more difficult, when administered to 60 ESL students than was a test with distractors written by two professional item writers using standard rules of item writing.

### 2.5.2 How to Improve Multiple-Choice Reading Tests

Aslanian (1985) provides seven suggestions on how to improve writing multiple-choice reading comprehension tests:

- (1) More authentic passages, with proper introductions, reasonable developments, and coherent conclusions ought to replace the choppy, inconsistent paragraphs currently used.
- (2) More developed, inferential, and paraphrased choices should replace the literal, obvious, and sometimes absurd choices that serve as distractors.
- (3) More variety of content and genre would make the test less culturally biased.
- (4) More of the questions should only be answerable after the student has read and comprehended the passage.
- (5) Very frequently, multiple-choice questions follow the order of the passage, thereby sending students to a limited part of the text for their answer. Chronological order encourages the isolation of bits of text rather than their meaningful synthesis ... If the questions were jumbled, an element of search would be incorporated into the process. Students would need some grasp of the text in its totality to be able to go to the relevant part.
- (6) Comprehension tests should contain questions on the figurative aspects of the passages and on evaluative or critical and affective comprehension. In other words, what the writer says should be as important as how and to what end he says it.
- (7) Most important, test scores should not be the only criterion by which to determine a student's

placement. They could be used in conjunction with such reading activities as oral reading and retelling, listening comprehension, summarizing and paraphrasing, and following written instructions. (p. 40-41)

### 2.5.3 An Alternative Approach to Multiple-Choice Reading Items

Munby (in Cohen, 1980) suggests another approach about multiple-choice reading comprehension items which involves oral discussion in small groups. In this approach, the distractors are partially correct, but there is only one completely correct answer. In order to select the best answer students need to use a particular "micro-skill" or a combination of micro-skills. Munby sees this approach as training students in problem solving through multiple-choice reading comprehension questions. First, the students read a passage silently and answer a series of multiple-choice questions. Then they get into groups of five or six to discuss multiple-choice questions and to come up with correct answers through majority vote. Then each group gives the reasons for rejecting the distractors in each question. Finally there is a class-level discussion of the items. Munby, in this way believes that students understand and come to conclusions by reasoning, rather than just guessing or recognizing the correct answer.

## 2.6 Retelling - a more direct way of testing reading

Retelling, as indicated earlier (Aslanian, 1985), is another method of testing reading. In the retelling procedure, after reading a passage the student is asked to repeat or retell as much as possible on what he has read and understood. While the student does the retelling, the teacher can check what the student recalls against a checklist. If the student has understood what he has read, he should be able to retrieve the information. In retelling there is individualized comprehension, that is, the student gives his own understanding of the passage. Retelling gives the reader's interpretation of the reading passage and from this interpretation it is easy to understand whether the student comprehended the passage or not.

Johnston (1983) says that the retelling procedure is quite time consuming as it requires individual testing, but with this testing technique we can learn something about how information is stored and organized in the memory, and what strategies are used in retrieving the stored memory. Johnston (1983, p. 55) believes that during the recall process " Patterns of intrusions, distortions and omissions may provide valuable information on specific influences of the individual's background knowledge".

Recall and retelling involve productive skills

rather than receptive skills. So failure to produce information could be due to production deficits. Alternatively, "schema selection or access may be blocked or inappropriate or it may be that the tester has read the passage, gives only a cursory protocol" (Johnston, 1983, p. 55).

## 2.6 Conclusions

Research has shown that multiple-choice tests are not favoured as tests of reading comprehension because of numerous drawbacks. These tests mostly depend on recognition and require the student to use his/her receptive skills, thus raising doubts as to what ability they are really measuring. Retelling, on the other hand, is solely based on productive skills, and requires the student to put his/her real language into use while completing the task.

From this research evidence it can be argued that the retelling procedure is a more authentic and reliable way of measuring reading comprehension as it is a more direct and communicative way of testing the reading abilities of students.



## CHAPTER III

### METHODOLOGY

#### 3.1 Introduction

Multiple-choice tests have a lot of advantages; in particular, they are objective and economical, that is, easy to score and time saving. Therefore, these tests have been widely used in EFL/ESL situations in the assessment of all language elements from phonological features like stress and intonation patterns to vocabulary, grammar and reading comprehension. However, in recent years, with the development of communicative tests, the validity of multiple-choice tests has been questioned, with multiple-choice being considered an indirect way of assessing language use. According to researchers like Oller (1979), Hughes (1989), Madsen (1983) and Aslanian (1985), the validity of multiple-choice tests as a testing technique for measuring reading comprehension is still debatable.

The concern of this study is to find out how well multiple-choice reading tests measure reading abilities of students. This is done by comparing this testing technique with another reading comprehension test technique "retelling", in which subjects are asked to give their individual oral account of the reading passage. The results obtained from the multiple-choice reading test and the retelling procedure are then

correlated with each other and with the teacher ratings and the BUSEL reading exam to determine which of these two reading tests highly correlate with the two outside measures.

### 3.2 Participants and Setting

#### 3.2.1 Subjects

The subjects of this study were Bilkent University School of English Language (BUSEL) first year intermediate level science students. Students who enter Bilkent University first take an English proficiency exam - the BUSEL Exemption Exam, and those who pass this exam become first year students. Those who are not successful in this exam take a Placement Test and from the scores they obtain in this test their proficiency levels are determined, and according to their levels they are then placed into beginner, elementary, intermediate or upper-intermediate classes. Therefore, all the BUSEL intermediate level subjects who took part in this study were considered to have similar proficiency levels of English.

A total of 18 subjects from an intermediate level science class participated in the study. Subjects were selected after a pilot study was done with one intermediate and two elementary classes. From the results of these 50 students an item analysis was done

(see Appendix B) and some changes were made in some of the distractors. From the results of this pilot study, the level of the subjects' was determined to be intermediate. Subjects were not informed about the nature of the study.

The study was conducted at BUSEL in the subjects' usual classrooms at their usual class hours with the intention of lowering stress and anxiety.

### 3.2.2 Experimenters

Since the retelling procedure requires individual interviews, it was conducted by the researcher and seven other Turkish teachers of English one of whom was the teacher of that class.

Teachers were given information on how to conduct the retelling procedure. The first thing the teachers were asked to do was to create a non-threatening atmosphere for the students like by asking their names. The next thing the teachers were asked to do was to tell the students that they could do the retelling in either language - English or Turkish. Then they were asked to tell the students to think about the passage they had just read, and ask them to recall and retell all the details they could remember about the passage. The teachers were also asked to inform the students that they were not going to be interrupted and they would be given

a second chance to think about the passage once again and add any additional points that came to their minds. For accuracy and efficiency, the teachers were given checklists to carry out the retelling procedure.

### 3.3 Materials

Materials used in this study included two reading comprehension tests: a multiple-choice reading comprehension test and a retelling checklist.

#### Instrument I

A reading passage of 469 words was chosen to be appropriate for intermediate level students (see Appendix C). This passage was used by Chavez-Oller et al. (1985) as a cloze task, and the original source was Praninskas (1959) (both cited in Jonz 1990). The passage was followed by eight multiple-choice comprehension questions (see Appendix D). The distractors of the multiple-choice test were prepared by taking Dobson's (1974) suggestion into account for preparing better distractors as mentioned in chapter II, that is, by choosing distractors from "high-frequency" incorrect student answers in completion items. To obtain incorrect student answers first, open-ended comprehension questions were prepared on the same reading passage and were given to the students of another intermediate level class. Then, from the usable wrong answers given to these questions, some

of the distractors for the multiple-choice questions were written.

### Instrument II

A checklist was prepared for the retelling procedure covering nearly all the important points of the reading passage. This was done by putting slashes in between each important piece of information in the passage. There were 40 total slashes and so 40 pieces of information to be given, but subjects were not expected to retell each and every one of these pieces as it was impossible to remember the whole text. The purpose for giving the checklist in a detailed form was to help the experimenters conducting the retelling procedure and also to give total credit to those subjects who remembered the passage with more details than the others (see Appendix E).

### 3.4 Two Outside Measures used as the Criteria for

#### Measuring Reading Ability

#### Teacher Ratings

Each class at BUSEL receives 20 hours of English instruction a week which is shared by two teachers. One of the teachers - the main teacher - teaches the class 12 hours a week, while the other meets the class for the remaining 8 hours. To increase the reliability of the

teacher ratings, both of the class teachers sharing the particular intermediate class were given a 1 to 5 Reading Ability Rating Scale (see Appendix F) and were asked to rate each students' reading ability on the scale by making joint decisions. These ratings were completed in the presence of the experimenter and the teachers rated students' reading abilities by taking their class performance into consideration, and without referring to any kind of exam or test result. The results of these ratings were then converted into percentages as all the other grades were in that form. The results of all the other tests were correlated with the teacher ratings as it was the primary criteria for measuring students' reading abilities.

#### BUSEL Reading Exam

The reading grades from the last BUSEL achievement exam were taken from the BUSEL administration. The reading part of the test was worth 30 points out of 100 and there were multiple-choice comprehension questions, true/false questions and reference questions. The reading scores were also converted into percentages (see Appendix A).

### 3.5 Procedures of Data Collection

Subjects were first given the reading passage with the multiple-choice questions with a time limit of 30 minutes. Instructions for the multiple-choice test were read aloud by the researcher, and they were also provided in writing. Subjects were told that they were going to be asked to do another activity with the same passage without telling the nature of the activity, that is, retelling.

After the completion of the first part of the test, the papers were collected, and eight subjects were randomly selected for the retelling, and the rest of the ten subjects were asked to wait outside the classroom. This was done to eliminate the possibility of students' interaction with each other and giving clues about the retelling procedure.

Each of the eight teachers worked with one subject. Before the procedure started subjects were informed that they were going to retell the story in their own words with as many details as they could remember, and that they would be given a second chance to think about the passage once again and give additional information if they remembered any. As subjects were not familiar with such an activity, they were also given the freedom of doing the retelling either in English or Turkish or using both if they liked so that they would feel more

comfortable. Subjects were given 10 minutes for this part of the test and the experimenters listened to their subjects' retelling without any interruption.

During the retelling procedure as the subjects retold the passage, the experimenters went through the checklist and put check marks against the items the subjects were able to remember. Once the subjects were through with the process of recalling, the experimenters asked the subjects if they had anything else to say or add to what they had already said. The same procedure was followed with the second half of the subjects. Subjects were graded on the number of items they could remember out of those 40 items on the checklist, and the subject with the highest score (28 out of 40) was taken as the top score and was considered as 100% and the scores were converted into percentages accordingly.

In the next chapter, the procedures used in collecting and analysing data will be discussed. An explanation of the statistical calculations and the results of the correlations between the different measures will be explained.



CHAPTER IV  
DATA ANALYSIS

4.1 Introduction

Recent research has questioned the validity of multiple-choice reading tests for measuring reading comprehension. The objective of this study was to correlate the multiple-choice reading test and the retelling procedure with the two outside measures, the teacher ratings and the BUSEL reading exam, and to see which of the two tests highly correlated with each other.

As mentioned earlier, the teacher ratings and the BUSEL reading exam were taken as the basis for a correlation. Students' reading abilities were rated by their two class teachers on a one to five scale, with 1 representing the poor reader and 5 representing the excellent reader (see Appendix F). In addition to the teacher rating, the results of the BUSEL reading exam were also taken as the second criteria for determining the reading ability of students.

Then, all these test results were correlated with each other, first with the Pearson Product Moment Correlation and then with the Tetrachoric Correlation, for the purpose of finding out which of the tests highly correlated with each other and especially to see whether the multiple-choice test results or the retelling

procedure results correlated higher with the teacher ratings and the BUSEL reading exam results.

An important clarification must be made concerning the interpretation of the data analysis. It is universally agreed that proving the validity of certain testing techniques is difficult if not impossible. In this study we are examining several techniques of testing reading comprehension and looking for high correlations between these techniques. If high correlations are found, however, we cannot say with complete certainty that any technique(s) is a valid measure of reading comprehension. What we can say if high correlations are found between any of these testing techniques is that they seem to be measuring the same things.

#### 4.2 Data Analysis

First, the Pearson Product Moment Correlation was used in calculating the correlations between tests. The Pearson Product Moment Correlation (PPMC) is "an index of the tendency for the scores of a group of examinees on one test to covary (that is, to differ from their respective mean in similar direction and magnitude) with the scores of the same group of examinees on another test" (Oller, 1979, p. 54).

Baker (1989) states that correlation coefficients help us discover what exactly a test is testing and if

different tests are measures of "the same thing" or not.

According to Baker (1989):

If the correlation coefficient between two sets of scores is sufficiently high, we may, in certain circumstances, conclude that roughly the same dimension of proficiency is being measured by each test. If, on the other hand, the correlation coefficient is low, we can conclude that the two aspects of proficiency which the test is measuring are relatively distinct. (p. 58)

Appendices G.1, G.2, G.3, G.4, G.5 and G.6, respectively, show the means, variances, standard deviations and the PPMC calculations between:

1. multiple-choice test and BUSEL reading exam
2. multiple-choice and teacher rating
3. BUSEL reading exam and retelling scores
4. retelling scores and teacher rating
5. BUSEL reading exam and teacher rating
6. multiple-choice and retelling scores.

Another statistical technique, the Tetrachoric Correlation, was then used to find out how well the multiple-choice test, the retelling scores and the BUSEL reading exam correlated with the teachers' ratings. According to Carroll and Hall (1985) teachers' judgements are a valid technique for measuring students' reading abilities.

#### 4.2.1 Pearson Product Moment Correlation

Table 4.1 shows all the possible PPM correlations between the multiple-choice test and the BUSEL reading exam, the multiple-choice test and the teacher ratings, the multiple-choice test and the retelling, the BUSEL reading exam and the retelling, the retelling and the teacher ratings, and finally the BUSEL reading exam and the teacher ratings.

Table 4.1

PPMC Calculations between the Four Tests

	M/C T.	BUSEL R.	Teacher R.	Retelling
Multiple-choice	1.0	0.34	0.15	0.41
BUSEL Reading	0.34	1.0	0.70	0.55
Teacher Rating	0.15	0.70	1.0	0.78
Retelling	0.41	0.55	0.78	1.0

As seen in the table the PPMC between the multiple-choice test and the BUSEL reading exam was found to be 0.34. Carroll and Hall (1985) state that for a correlation to be significant, it should be at least 0.54

with 20 testees, 0.35 with 50 testees and 0.25 with 100 testees. Therefore, a correlation of 0.34 with 18 testees is very low, and since correlation is a measure of the relatedness of two variables, this result implies that the two tests, multiple-choice and the BUSEL reading exam, do not measure the same things.

The correlation between the multiple-choice test and the teachers' rating was found to be  $r = 0.15$  which is an even lower correlation than the one above. According to this result we can say that the two tests do not correlate with each other, i.e., they measure two different things rather than measuring the same variables.

In the same table it can be seen that there was a positive correlation  $r = 0.55$  (common variance = 0.3025) between the retelling procedure and the BUSEL reading exam. This is significant at the .05 level for a non-directional (2-tailed) test ( $df = 17$ ). Thus, we can conclude that these two tests have a positive correlation between them and that the two tests are measures of the same thing.

A higher correlation can be seen between the retelling scores and the teachers' rating. The correlation was found to be  $r = 0.78$  (common variance = 0.6084) which is significant at the .001 level ( $df = 17$ ). This high correlation indicates that at face value these

two tests measure the same things or that what can be known from one test can be known equally well from the other.

The correlation between the BUSEL reading exam and the teachers' rating was done to find out how reliable the reading exam results were and whether they reflected teacher ratings. The high correlation of  $r = 0.70$  helps us confirm that the results of the reading exam were sound reflections of the teachers' rating.

The correlation between the multiple-choice test and the retelling was also calculated although the PPMC is a correlation calculation for two independent sets of tests. These two tests were based on the same single reading passage and so might not be considered as two independent tests. However, the procedures carried out and the scores obtained for each test were totally different from each other. The PPMC result of  $r = 0.41$  show a dissimilarity between the two tests.

#### 4.2.2 Tetrachoric Correlation

As another way to find out how well students' reading abilities were reflected by the three tests, multiple-choice, retelling and BUSEL reading exam, the results of these tests were correlated by a technique called the Tetrachoric Correlation. Carroll and Hall (1985) define this technique as the agreement between the

test results and the teachers' judgements of the competence of the students. This technique is based on the assumption that the students rated as competent by the teacher should also do well on the test, and those rated as sub-standard should do poorly on it. From this statement it is once again seen that Carroll and Hall value teacher judgements to a great extent and consider them as sound and reliable reflections of students' abilities. Although teachers' ratings and the test results are expected to agree, there is the chance that some students judged to be competent by the teacher will fail and some judged sub-standard will pass, but these are hoped to be a fairly small proportion of the total.

In order to carry out the tetrachoric correlation the test results (multiple-choice, retelling and reading exam) had to be given pass and fail marks, and the teachers' ratings of the students had to be expressed as competent or sub-standard (i.e good readers as competent and poor readers as sub-standard). These ratings are:

Competent		Sub-standard
Pass (a)		(b) Pass
Competent		Sub-standard
Fail (c)		(d) Fail

To classify students as being competent or sub-standard two teachers sharing the particular intermediate

level class were asked to rate students' reading abilities on a 1 to 5 scale (see Appendix F). Each point on the scale was converted into percentages because the other test results were all expressed in percentages. Therefore, point 1 on the scale was equivalent to 20%, point 2 to 40%, point 3 to 60%, point 4 to 80% and point 5 corresponded to 100%. Three points on the scale was considered as average (competent), and the students who got 3 points were accepted as competent students. Those who were rated between 1 and 3 were considered to be sub-standard, while others rated between 3 and 5 on the scale were considered to be competent readers. For example, a student given 2 on the scale by his/her teachers was taken to be a poor reader (sub-standard) because 2 corresponds to 40 out of 100, whereas, a student who was given 4 on the scale, which corresponds to 80 out of 100, was considered to be a good (competent) reader.

Similarly, to decide on a pass or fail mark for students' test results, the percentiles of the scores of each student were calculated for each of the three tests as Baker (1989) suggests using percentile scores for assigning pass and fail marks (see Appendix A for raw scores and percentile scores of each test). Percentiles show where a testee stands in relation to other testees. After the percentiles scores were calculated, two axes were drawn with 100 divisions vertically representing the



percentile scores, and 100 divisions were drawn horizontally representing the raw exam results. Starting with the lowest grade and working up to the highest, the percentile scores of each test were plotted against their raw scores on the horizontal axis. To determine the pass mark a line was drawn across from the 50th percentile of each of the separate tests, and the pass mark was obtained by reading off the horizontal axis at the point where the line intersected the curve (Baker, 1989).

Appendices H.1, H.2 and H.3 illustrate the percentile and raw scores of the three tests with the pass mark for each test. As seen on the graphs, the pass mark assigned for multiple-choice test was 72, for retelling it was 63, and for the BUSEL reading exam it was 71.

Tetrachoric correlation between multiple-choice test results and teachers' ratings are shown in Table 4.2 (see Appendix A for each testees test results and percentile scores).

**Table 4.2**  
Tetrachoric Correlation Between  
the Multiple-Choice Test and the Teachers' Rating

		Teacher Judgement	
		Competent	Sub-standard
Test Results	Pass	10 (a)	(b) 2
	Fail	5 (c)	(d) 1

$$r_{tet} = \cosine \left( \frac{bc}{ad + bc} \right) 180$$

$$r_{tet} = 0$$

According to the teachers' judgement, 15 students were judged to be competent and 3 students were judged to be sub-standard. Although the teachers had judged 15 students as competent only 10 of these competent students passed the multiple-choice test and the other 5 of them failed. On the other hand, of the 3 sub-standard students only 1 failed and the other 2 passed the test. When the tetrachoric correlation was calculated the result was  $r_{tet} = 0.00$ , meaning there was no correlation whatsoever between the multiple-choice test results and the teachers' judgement. As the teachers' judgement was taken to be a reliable measure for representing students' true reading abilities, it can be said that multiple-choice test scores do not reflect students' reading ability. When the PPMC and Tetrachoric correlations for

multiple-choice and teachers' judgements are compared they are  $r = 0.15$  and  $r_{tet} = 0.00$  respectively. Both low correlations once more strengthen the implication that multiple-choice test results are a measure of a different ability.

**Table 4.3**  
Tetrachoric Correlation Between  
Retelling and the Teachers' Rating

		Teacher Judgement	
		Competent	Substandard
Test	Pass	9 (a)	(b) 1
Results	Fail	6 (c)	(d) 2

$$r_{tet} = \cosine \left( \frac{bc}{ad + bc} \right) 180$$

$$r_{tet} = 0.41$$

As shown in table 4.3, the correlation between retelling scores and teachers' judgement gave us  $r_{tet} = 0.41$ . Nine out of 15 competent students passed while the remaining 6 failed. One sub-standard student passed and the other 2 failed. The reason why this correlation between retelling and the teachers' rating is low might be due to the fact the criteria for a good reader at BUSEL is considered a high performance on an open-ended reading comprehension test or a multiple-choice test. The class teachers' judgement of good readers is probably

based on the students' performance in reading comprehension activities and tests which require the use of receptive rather than productive skills, and therefore, the correlation between the two measures do not show a very close similarity. Although this correlation is not as high as expected by comparing it with the high PPMC correlation of  $r = 0.78$  for the same sets of tests, there is still a positive correlation between the two test results. In other words, retelling reflects students' reading abilities to a greater extent than the multiple-choice test does.

The tetrachoric correlation between the BUSEL reading exam and teachers' judgement show that out of 15 10 competent students passed the reading test, whereas, the other 5 students failed (see Table 4.4). On the other hand, the 3 sub-standard students all failed which shows that there is a match between the teacher judgements of the students and the grades of the BUSEL reading exam.

**Table 4.4**  
Tetrachoric Correlation Between  
the BUSEL Reading Exam and the Teachers' Rating  
Teacher Judgement

		Competent	Sub-standard
Test Results	Pass	10 (a)	(b) 0
	Fail	5 (c)	(d) 3

$$r_{tet} = \cosine \left( \frac{bc}{ad + bc} \right) 180$$

$$r_{tet} = 1$$

Table 4.4 shows a perfect correlation of  $r_{tet} = 1.0$  between the reading exam and teachers' judgement. Ten out of 15 competent students passed the test and only 5 failed. On the other hand, all 3 of the sub-standard students failed the test. This perfect correlation show that these tests totally agree on what they are measuring. The PPMC for the same tests was  $r = 0.70$ , another high correlation approving the similarity of the two measures. In other words, these correlations are significantly high enough to say that the BUSEL reading exam and the teacher ratings are measuring the same abilities.

#### 4.3 Results

From the results of the PPMC and Tetrachoric correlations done between the four kinds of rating of

students' reading ability, the retelling scores and the BUSEL reading exam gave the highest correlations. We can thus conclude that retelling scores reflect students reading abilities better than multiple-choice tests. In other words, from the results of the retelling correlations it can be said that retelling and teacher judgements are measuring two similar abilities and two similar aspects of proficiency. The same is true for the BUSEL reading exam as the results of this exam highly correlated with the teachers' judgements. The results of the teachers' judgement and the reading exam perfectly matched and brought about a full correlation of  $r_{tet} = 1.00$ . This clearly shows that these two tests are measuring the same things and when the retelling scores are compared to these two tests, the result is a high enough correlation to accept it as a measure of the same variable.

The results of the correlations done between multiple-choice and teacher rating, and multiple-choice and the BUSEL reading exam were found to be either very low or zero, showing that there is no similarity in whatever they are measuring.

## CHAPTER V

### CONCLUSIONS

#### 5.1 Summary of the Study

The focus of this study was on the validity of multiple-choice reading comprehension tests and the retelling testing procedure among other measures used for assessing reading comprehension abilities. These two tests were correlated with each other and with two outside measures, the class teachers' judgements of the subjects' reading abilities and the BUSEL reading exam results.

As mentioned in the previous chapters the validity of multiple-choice reading tests is debatable. These tests have a number of disadvantages but are widely used in EFL/ESL situations testing vocabulary, grammar and reading because they are objective, and very easy and economical to score. On the other hand, there are a lot of drawbacks with multiple-choice tests, especially multiple-choice reading comprehension tests. Multiple-choice tests require the testee to read and answer questions, but they usually leave open the question of what the test is a test of (Oller, 1979). As these tests are recognition type of tests they are indirect ways of testing reading comprehension, therefore, it is usually very difficult to understand whether a testee really

comprehended a passage only by looking at the score he obtained from a multiple-choice reading test.

Another disadvantage that multiple-choice tests have is that it is very difficult to prepare multiple-choice items and distractors. While preparing these tests there are a number of principles that should be taken into consideration, the most important being the relevance of the test items to the skill that is to be measured. Suggestions for preparing better multiple-choice reading tests were dealt with in Chapter II.

The retelling procedure, on the other hand, is considered to be a more direct and communicative way of testing reading comprehension since there is a similarity between the test performance and the criterion performance. According to Baker (1979), the test performance is what the testee has to do during the test, and the criterion performance is what the testee would have to do in a real test situation. Therefore, retelling, being a direct way of testing reading, seems to be a more valid technique than multiple-choice because direct tests give a better idea about the testees real language use.

To find out what these two tests measured, multiple-choice and retelling, they were compared and correlated with the two outside measures, teacher ratings and the BUSEL reading exam. A reading passage was chosen for



intermediate level BUSEL students. Eight multiple-choice questions were prepared taking into account the principles and suggestions for preparing better items and a checklist was prepared for the retelling procedure, and these two tests were given to 18 BUSEL intermediate level science students.

The results obtained from these two tests were then correlated with two outside measures, teachers' rating and the BUSEL reading exam.

## 5.2 Assessment of the Study

The Pearson Product Moment Correlation and the Tetrachoric Correlations were used to find out how well the multiple-choice test and the retelling results correlated with teacher ratings and the BUSEL reading exam (see Chapter IV, section 4.2 Data Analysis).

The Pearson Product Moment Correlation was calculated between the multiple-choice test and the BUSEL reading exam, the multiple-choice test and the teacher ratings, the multiple-choice test and retelling, retelling and the BUSEL reading exam, retelling and the teacher ratings, and the BUSEL reading exam and the teacher ratings (see Table 4.1 in Chapter IV). The other statistical technique, the Tetrachoric Correlation, was then used to find out how well the multiple-choice test, the retelling scores and the BUSEL reading exam

correlated with the teachers' rating (see Tables 4.2, 4.3, 4.4 in Chapter IV).

From the results of these two different correlation calculations it was found that the PPMC between the multiple-choice test and the BUSEL reading exam was  $r = 0.34$ , and the PPMC between the multiple-choice test and the teachers' rating was  $r = 0.15$ . The Tetrachoric Correlation between the multiple-choice test and the teachers' rating was found to be  $r = 0$ . By looking at these considerably low correlations it can be concluded that the multiple-choice test is measuring something different than the two outside measures, teacher ratings and the BUSEL reading exam.

The PPMC done between the retelling scores and the BUSEL reading exam was  $r = 0.55$  and the PPMC between the retelling scores and the teachers' rating was  $r = 0.78$ . The result of the Tetrachoric Correlation between the retelling scores and the teachers' rating was found to be  $r_{tet} = 0.41$ . These relatively high correlations demonstrate that there is a considerable similarity between the retelling procedure and the teachers' rating and the BUSEL reading exam. Thus, it can be concluded that the retelling scores reflected students' reading abilities better than the multiple-choice test.

The high PPMC of  $r = 0.70$  between the BUSEL reading exam and the teachers' rating, and the perfect

Tetrachoric Correlation of  $r_{tet} = 1.00$  between the same tests once more indicate the reliability of these results and show that they are measures of the same things. And as mentioned above, when the retelling scores are based on these two outside measures, we obtain a correlation which is high enough to claim that they are all measures of the same variable.

These findings prove the experimental hypothesis (b) of the study to be true: " There is a positive correlation between the retelling procedure and the teachers' rating and the BUSEL reading exam". The findings for the multiple-choice test, on the other hand reject the experimental hypothesis but agree with the null hypothesis: " There is no correlation between the multiple-choice test and the teachers' rating and the BUSEL reading exam". These results also show an agreement with what the literature says.

### 5.3 Pedagogical Implications

Multiple-choice reading tests are popularly and widely used but most of the time without questioning what these tests are really measuring. These tests are preferred especially when large numbers of testees are being tested because the scoring is relatively quick and easy. The primary purpose when giving a multiple-choice reading test to a certain group of testees should be to

make sure that the test is measuring what it sets out to measure. In other words, the test should be congruent with the skill that is being measured and should reflect students' abilities in that particular skill.

In schools and universities where the medium of instruction is English, greater emphasis should be given in the preparation of multiple-choice tests. The construction of multiple-choice tests is quite complicated, difficult and time consuming, and therefore, should be approached with great care. However, it is possible to prepare more reliable and valid multiple-choice items by taking the suggestions of researchers like Dobson (1974), Aslanian (1985), Munby (1980), Oller (1979) and others into consideration.

#### 5.4 Implications for Future Research

In this study the validity of multiple-choice reading tests and of the retelling procedure was compared and correlated with two outside criteria, teachers' rating and the BUSEL reading exam. The same reading passage and the same subjects were used in carrying out both the multiple-choice test and the retelling procedure. Actually it is not a natural procedure to complete a multiple-choice reading test, and then right afterwards ask the testee to retell the same passage. The testees are usually not exposed to such a procedure

in their normal classroom activities, but due to time constraints and limitations, like finding testees with the same level of proficiency and finding two very similar reading passages, the study had to be carried out in this way.

Another possible approach can be followed for the correlation of multiple-choice reading tests and the retelling procedure by using two different reading passages that are similar in content, structure and level. For one of the reading passages multiple-choice questions can be prepared, and for the other a checklist to conduct the retelling procedure. Then the two different tests can be given to the same group of subjects all having the same proficiency level of English. First, they can be asked to complete the passage with the multiple-choice questions, and then later they can be given the other reading passage and this time asked to read and retell the passage. As a moderator variable the retelling procedure can be conducted by allowing the subjects to use only the target language as the original text was given in English, and to see whether the superior retelling of some subjects was a function of the language medium that was used. These changes in the approach of the study would allow us to have two independent tests and test results, and a correlational study between these two independent tests

would help us see which of the two tests reflected students' reading abilities better.

## BIBLIOGRAPHY

- Aslanian, Y. (1985). What's wrong with multiple-choice tests? Selected Papers from the 1985 Annual Conference of the CUNY ESL Council. New York: Instructional Resource Center.
- Baker, D. (1989). Language testing: A critical survey and practical guide. London: Edward Arnold.
- Bensoussan, M. (1984). A comparison of cloze and multiple-choice reading comprehension tests of English as a Foreign Language. Language Testing, 1 (1), 101-104.
- Carroll, B. J. and Hall, P. J. (1985). Make your own language tests: A practical guide to writing language performance tests. Pergamon Press.
- Cohen, A. D. (1980). Testing language ability in the classroom. Rowley, Massachusetts: Newbury House Publishers.
- Galloway, V. (1987). From defining to developing proficiency: A look at the decisions. In H. Byrnes and M. Canale (Eds.), Defining and Developing Proficiency (pp. 75-96). Illinois: National textbook Company.
- Heaton, J. B. (1988). Writing English language tests: A practical guide for teachers of English as a Second or Foreign Language. London: English Language Book Society, Longman.
- Hughes, A. (1989). Testing for language teachers. Cambridge: Cambridge University Press.
- Johnston, P. H. (1983). Reading comprehension assessment: A cognitive basis. Newark, Delaware: International Reading Association.
- Jonz, J. (1990). Another turn in the conversation: What does cloze measure? TESOL Quarterly, 24 (1), 78-79.
- Madsen, H. S. (1983). Techniques in testing. New York and Oxford: Oxford University Press.
- Oller, J. W. (1979). Language tests at school: A pragmatic approach. London: Longman.

- Riley, P.M. (1977). Specific reading skills for secondary and post secondary ESL learners. TEFL/TESL Newsletter, 3, 1-6.
- Singer, H. and Ruddell, R.B. (Eds.). (1985). Theoretical models and processes of reading. Newark, Delaware: International Reading Association.
- Valette, R. M. (1977). Modern language testing. USA: Harcourt Brace Jovanovich, Inc.
- Wood, N. V. (1987). Standardized reading tests and the postsecondary reading curriculum. (Report No. CS 008 840). Paper presented at the Annual Meeting of the Western College Reading and Learning Association. (ERIC Document Reproduction Service No. ED 283 128).



## APPENDIX A

## RAW SCORES and PERCENTILES

TESTEE	M/C MARK/Perc.	RETEL. MARK/Perc.	READING EXAM/Perc.	TEACHER RATING
1	100/1.00	100/1.00	90/0.94	100
2	100/1.00	96/0.94	73/0.61	100
3	100/1.00	89/0.83	73/0.61	100
4	100/1.00	64/0.61	60/0.33	40
5	100/1.00	82/0.78	93/1.00	100
6	88/0.72	96/.094	90/0.94	100
7	88/0.72	57/0.39	60/0.33	80
8	88/0.72	64/0.61	83/0.83	80
9	75/0.56	39/0.22	53/0.17	40
10	75/0.56	61/0.44	60/0.33	60
11	75/0.56	39/0.22	70/0.44	80
12	75/0.56	43/0.33	73/0.61	60
13	63/0.33	32/0.11	50/0.11	40
14	63/0.33	29/0.06	63/0.39	60
15	63/0.33	43/0.33	77/0.78	60
16	50/0.17	64/0.61	77/0.78	100
17	50/0.17	79/0.72	50/0.11	80
18	38/0.06	79/0.72	77/0.78	100

APPENDIX B  
Item Analysis of the Pilot Test  
with 50 subjects

Item	Upper Group	Lower Group	Diff. Ind.	Disc. Ind.
1	14	10	48.00	0.16
2	8	3	22.00	0.20
3	16	9	50.00	0.28
4	19	10	58.00	0.36
5	7	1	16.00	0.32
6	19	9	56.00	0.40
7	20	13	66.00	0.28
8	16	11	54.00	0.20

APPENDIX C  
READING PASSAGE  
LEAVING HOME

Joe is a freshman, and he is having all the problems that most freshmen have. As a matter of fact, his problems started before he even left home. He had to do a lot of things he didn't like to do just because he was going to go away to college. He had his eyes examined and he had his cavities filled, although he hates to go to a dentist, and he got his watch fixed by a neighborhood jeweler. He cleaned his room, packed his old books, and left aside the ones that he thought would be useful for him at college. Then, at his mother's suggestion, he had his father's tailor measure him for a suit. He didn't have a suit made, though, because his father wouldn't let him order one. "You're still growing, son," he said. "You're growing so fast that you'd outgrow a suit in no time. Buy yourself a pair of slacks and a sports jacket. Klein's has such a large selection that I'm sure you will find something you like there." Joe's father always suggested Klein's for clothes.

Joe went to Klein's in order to please his father, but he didn't find anything that he liked there so he went to another store and bought a pair of blue slacks

and a sports jacket that went well with it. He took them out of the box as soon as he got home so that his father wouldn't notice where they came from.

Joe phoned his best friends, gave them his address, and they all wished him good luck in his new life, and promised him that they would write to him and maybe even visit him. When he was all ready to leave for school, his mother wanted him to visit all his relatives. "What do you want me to do that for?" he asked, and she answered, "To say good-bye like you did with your friends." She made him go to see his cousins in Bellevue and his Uncle Ned in Plaintown and his Great-Aunt Lizzie who lives in the southern part of the state. He didn't want to visit all those people, but he did it anyway because of his mother's insistence.

On the day that he left for college, his sister helped him pack his clothes and books. She let him borrow her suitcase because he didn't have one of his own. When everything was all ready, he got his father to drive him to the station, and the whole family went along. Of course, his mother insisted on kissing him good-bye in spite of his embarrassment. As soon as the train pulled into the station, Joe jumped on and hurriedly found his seat. By the time it pulled out, he was already contemplating his new life away from home.

## APPENDIX D

## MULTIPLE-CHOICE QUESTIONS

Circle the letter of the best answer.

- 1) What is the main idea of this passage?
  - a. Joe had a lot of problems at home.
  - \* b. Joe had to do many things before leaving for college.
  - c. Freshmen have a lot of problems all the time.
  - d. Leaving home is usually hard but fun.
  
- 2) What did he do with his old books?
  - a. He took all of them with him to college.
  - b. He packed them all up, and left them at home.
  - c. He packed and left them all aside.
  - \* d. He left some at home, and took the rest to college.
  
- 3) What did Joe's mother suggest that he do? She suggested that he...
  - a. buy clothes for college.
  - \* b. go to his father's tailor.
  - c. pack his books and clothes.
  - d. measure himself.
  
- 4) What does Joe's father mean by saying "you'd outgrow a suit in no time?"
  - a. You don't need to have a suit for the time being.
  - \* b. You are still growing and so don't have a suit

made.

c. You are too young to have a suit made.

d. You don't have enough time to have a suit made.

5) Why didn't Joe want his father to see the box?

Because...

\* a. he didn't want his father to understand where he shopped.

b. he bought things that his father wouldn't like.

c. he didn't want his father to see his new clothes.

d. he bought too many things for college.

6) What did Joe's friends promise?

a. They would send their addresses.

b. They would definitely visit him.

\* c. They would write him.

d. They would phone him.

7) Why did Joe visit his relatives? Because...

\* a. his mother insisted that he visit them.

b. he didn't want to phone them.

c. he wanted to say good-bye to them.

d. he was going away forever.

8) Why was Joe embarrassed? Because...

a. he left his family and friends.

\* b. his mother kissed him.

c. he didn't want to leave home.

d. his sister gave him her suitcase.

\* correct answer

## APPENDIX E

## CHECKLIST

Joe is a freshman / he has problems that most freshmen have / his problems started before he left home / he had to do a lot of things / he didn't like to do them / he had his eyes examined / had his cavities filled / he hates to go to a dentist / he got his watch fixed by a jeweler / he cleaned his room / packed his old books / left aside the ones that would be useful for him at college / his mother suggested that his father's tailor measure him for a suit / he didn't have a suit made / his father wouldn't let him order one / " You're still growing and you'd outgrow a suit in no time" he said / buy yourself slacks and a sports jacket / Klein's has a large selection / you will find something you like there. Joe went to Klein's / he didn't find anything he liked / he went to another store / bought blue slacks and a sports jacket / took them out of the box at home / didn't want his father to notice where they were from.

Joe phoned his friends / gave them his address / they wished him good luck / promised him that they would write to him, maybe even visit him / his mother wanted him to visit his relatives to say good-bye / he went to his cousins, uncle and Great-Aunt / he didn't want to visit but his mother insisted.

His sister helped him pack his clothes and books / she let him borrow her suitcase / his father drove him to the station / the whole family went along / his mother insisted on kissing him / he was embarrassed / Joe jumped on the train and found his seat / he was contemplating his new life away from home.

## APPENDIX F

## Reading Ability Rating Scale

	1	2	3	4	5
Points	Poor Reader		Average Reader		Excellent Reader
%	20	40	60	80	100



## APPENDIX G.1

PPM CORRELATION BETWEEN MULTIPLE-CHOICE TEST (X) AND BUSSEL READING EXAM (Y)

NO	X	Y	X-M	Y-M	(X-M)*(Y-M)	(X-M)^2	(Y-M)^2
1	100	90	22.72	19.33	439.30	516.30	373.78
2	100	73	22.72	2.33	53.02	516.30	5.44
3	100	73	22.72	2.33	53.02	516.30	5.44
4	100	60	22.72	-10.67	-242.37	516.30	113.78
5	100	93	22.72	22.33	507.46	516.30	498.78
6	88	90	10.72	19.33	207.30	114.97	373.78
7	88	60	10.72	-10.67	-114.37	114.97	113.78
8	88	83	10.72	12.33	132.24	114.97	152.11
9	75	53	-2.28	-17.67	40.24	5.19	312.11
10	75	60	-2.28	-10.67	24.30	5.19	113.78
11	75	70	-2.28	-0.67	1.52	5.19	0.44
12	75	73	-2.28	2.33	-5.31	5.19	5.44
13	63	50	-14.28	-20.67	295.07	203.85	427.11
14	63	63	-14.28	-7.67	109.46	203.85	58.78
15	63	77	-14.28	6.33	-90.43	203.85	40.11
16	50	77	-27.28	6.33	-172.76	744.08	40.11
17	50	50	-27.28	-20.67	563.74	744.08	427.11
18	38	77	-39.28	6.33	-248.76	1542.74	40.11
SUM	1391	1272			1552.67	6589.61	3102.00
MEAN	77.28	70.67					
				VARIANCE		366.09	172.33
				STD. DEV.		19.13	13.13
				PPMC	0.34		

## APPENDIX G.2

PPM CORRELATION BETWEEN MULTIPLE-CHOICE TEST (X) AND TEACHER RATING (Y)

NO	X	Y	X-M	Y-M	(X-M)*(Y-M)	(X-M)^2	(Y-M)^2
1	100	100	22.72	23.33	530.19	516.30	544.44
2	100	100	22.72	23.33	530.19	516.30	544.44
3	100	100	22.72	23.33	530.19	516.30	544.44
4	100	40	22.72	-36.67	-833.15	516.30	1344.44
5	100	100	22.72	23.33	530.19	516.30	544.44
6	88	100	10.72	23.33	250.19	114.97	544.44
7	88	80	10.72	3.33	35.74	114.97	11.11
8	88	80	10.72	3.33	35.74	114.97	11.11
9	75	40	-2.28	-36.67	83.52	5.19	1344.44
10	75	60	-2.28	-16.67	37.96	5.19	277.78
11	75	80	-2.28	3.33	-7.59	5.19	11.11
12	75	60	-2.28	-16.67	37.96	5.19	277.78
13	63	40	-14.28	-36.67	523.52	203.85	1344.44
14	63	60	-14.28	-16.67	237.96	203.85	277.78
15	63	60	-14.28	-16.67	237.96	203.85	277.78
16	50	100	-27.28	23.33	-636.48	744.08	544.44
17	50	80	-27.28	3.33	-90.93	744.08	11.11
18	38	100	-39.28	23.33	-916.48	1542.74	544.44
SUM	1391	1380			1116.67	6589.61	9000.00
MEAN	77.28	76.67					
				VARIANCE		366.09	500.00
				STD. DEV.		19.13	22.36
				PPMC	0.15		

## APPENDIX G.3

PPM CORRELATION BETWEEN BUSEL READING EXAM (X) AND RETELLING (Y)

NO	X	Y	X-M	Y-M	(X-M)*(Y-M)	(X-M)^2	(Y-M)^2
1	90	100	19.33	35.78	691.70	373.78	1280.05
2	73	96	2.33	31.78	74.15	5.44	1009.83
3	73	89	2.33	24.78	57.81	5.44	613.94
4	60	64	-10.67	-0.22	2.37	113.78	0.05
5	93	82	22.33	17.78	397.04	498.78	316.05
6	90	96	19.33	31.78	614.37	373.78	1009.83
7	60	57	-10.67	-7.22	77.04	113.78	52.16
8	83	64	12.33	-0.22	-2.74	152.11	0.05
9	53	39	-17.67	-25.22	445.59	312.11	636.16
10	60	61	-10.67	-3.22	34.37	113.78	10.38
11	70	39	-0.67	-25.22	16.81	0.44	636.16
12	73	43	2.33	-21.22	-49.52	5.44	450.38
13	50	32	-20.67	-32.22	665.93	427.11	1038.27
14	63	29	-7.67	-35.22	270.04	58.78	1240.60
15	77	43	6.33	-21.22	-134.41	40.11	450.38
16	77	64	6.33	-0.22	-1.41	40.11	0.05
17	50	79	-20.67	14.78	-305.41	427.11	218.38
18	77	79	6.33	14.78	93.59	40.11	218.38
SUM	1272	1156			2947.33	3102.00	9181.11
MEAN	70.67	64.22					
				VARIANCE		172.33	510.06
				STD. DEV.		13.13	22.58
				PPMC	0.55		

## APPENDIX G.4

PPM CORRELATION BETWEEN RETELLING SCORE (X) AND TEACHER RATING (Y)

NO	X	Y	X-M	Y-M	(X-M)*(Y-M)	(X-M)^2	(Y-M)^2
1	100	100	35.78	23.33	834.81	1280.05	544.44
2	96	100	31.78	23.33	741.48	1009.83	544.44
3	89	100	24.78	23.33	578.15	613.94	544.44
4	64	40	-0.22	-36.67	8.15	0.05	1344.44
5	82	100	17.78	23.33	414.81	316.05	544.44
6	96	100	31.78	23.33	741.48	1009.83	544.44
7	57	80	-7.22	3.33	-24.07	52.16	11.11
8	64	80	-0.22	3.33	-0.74	0.05	11.11
9	39	40	-25.22	-36.67	924.81	636.16	1344.44
10	61	60	-3.22	-16.67	53.70	10.38	277.78
11	39	80	-25.22	3.33	-84.07	636.16	11.11
12	43	60	-21.22	-16.67	353.70	450.38	277.78
13	32	40	-32.22	-36.67	1181.48	1038.27	1344.44
14	29	60	-35.22	-16.67	587.04	1240.60	277.78
15	43	60	-21.22	-16.67	353.70	450.38	277.78
16	64	100	-0.22	23.33	-5.19	0.05	544.44
17	79	80	14.78	3.33	49.26	218.38	11.11
18	79	100	14.78	23.33	344.81	218.38	544.44
SUM	1156	1380			7053.33	9181.11	9000.00
MEAN	64.22	76.67					
			VARIANCE			510.06	500.00
			STD. DEV.			22.58	22.36
			PPMC		0.78		

## APPENDIX G.5

PPM CORRELATION BETWEEN BUSEL READING EXAM (X) AND TEACHER RATING (Y)

NO	X	Y	X-M	Y-M	(X-M)*(Y-M)	(X-M)^2	(Y-M)^2
1	90	100	19.33	23.33	451.11	373.78	544.44
2	73	100	2.33	23.33	54.44	5.44	544.44
3	73	100	2.33	23.33	54.44	5.44	544.44
4	60	40	-10.67	-36.67	391.11	113.78	1344.44
5	93	100	22.33	23.33	521.11	498.78	544.44
6	90	100	19.33	23.33	451.11	373.78	544.44
7	60	80	-10.67	3.33	-35.56	113.78	11.11
8	83	80	12.33	3.33	41.11	152.11	11.11
9	53	40	-17.67	-36.67	647.78	312.11	1344.44
10	60	60	-10.67	-16.67	177.78	113.78	277.78
11	70	80	-0.67	3.33	-2.22	0.44	11.11
12	73	60	2.33	-16.67	-38.89	5.44	277.78
13	50	40	-20.67	-36.67	757.78	427.11	1344.44
14	63	60	-7.67	-16.67	127.78	58.78	277.78
15	77	60	6.33	-16.67	-105.56	40.11	277.78
16	77	100	6.33	23.33	147.78	40.11	544.44
17	50	80	-20.67	3.33	-68.89	427.11	11.11
18	77	100	6.33	23.33	147.78	40.11	544.44
SUM	1272	1380			3720.00	3102.00	9000.00
MEAN	70.67	76.67					
				VARIANCE		172.33	500.00
				STD. DEV.		13.13	22.36
				PPMC	0.70		

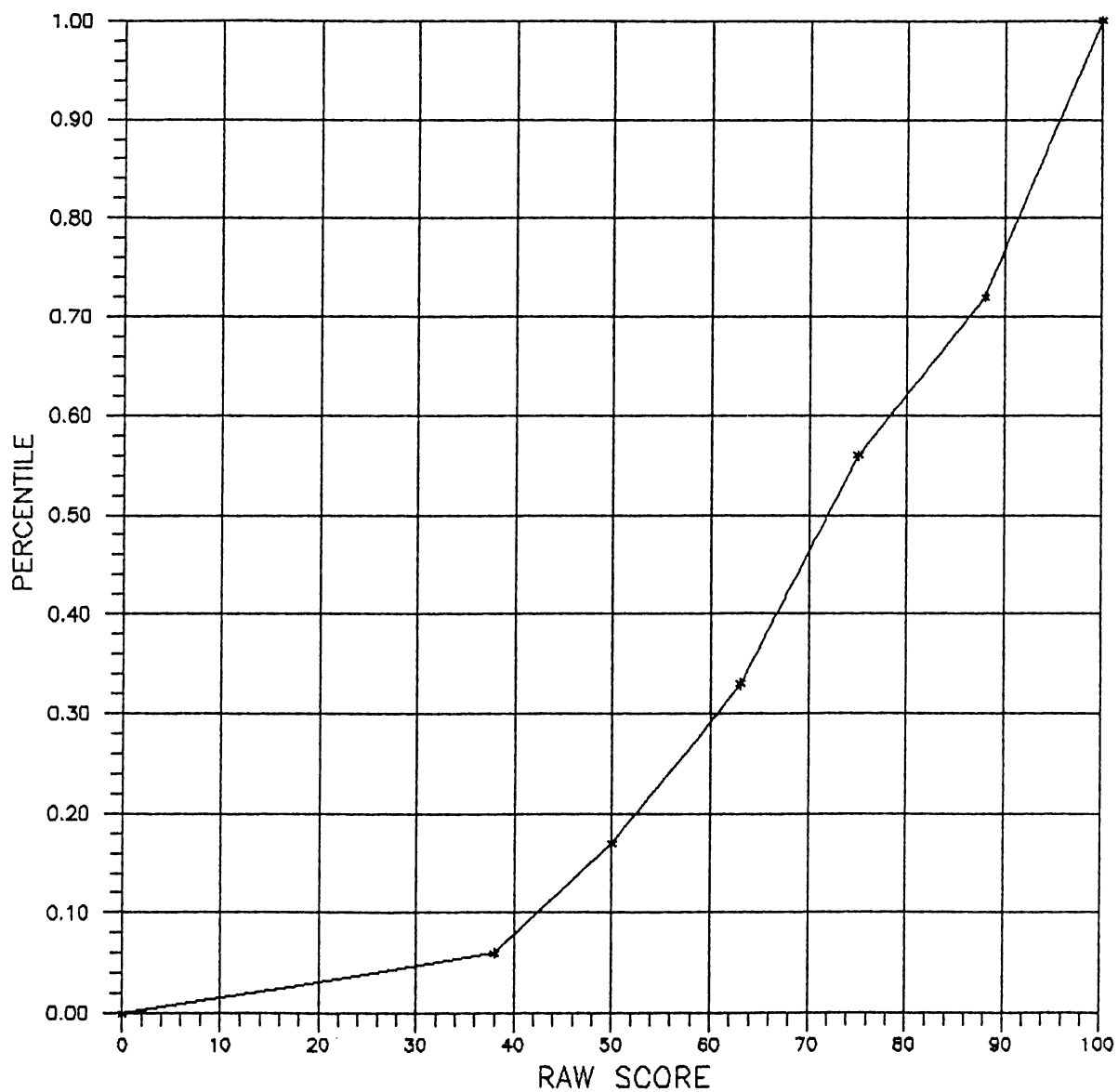
## APPENDIX G.6

PPM CORRELATION BETWEEN MULTIPLE-CHOICE TEST (X) AND RETELLING SCORE (Y)

NO	X	Y	X-M	Y-M	(X-M)*(Y-M)	(X-M)^2	(Y-M)^2
1	100	100	22.72	35.78	812.95	516.30	1280.05
2	100	96	22.72	31.78	722.06	516.30	1009.83
3	100	89	22.72	24.78	563.01	516.30	613.94
4	100	64	22.72	-0.22	-5.05	516.30	0.05
5	100	82	22.72	17.78	403.95	516.30	316.05
6	88	96	10.72	31.78	340.73	114.97	1009.83
7	88	57	10.72	-7.22	-77.44	114.97	52.16
8	88	64	10.72	-0.22	-2.38	114.97	0.05
9	75	39	-2.28	-25.22	57.45	5.19	636.16
10	75	61	-2.28	-3.22	7.34	5.19	10.38
11	75	39	-2.28	-25.22	57.45	5.19	636.16
12	75	43	-2.28	-21.22	48.34	5.19	450.38
13	63	32	-14.28	-32.22	460.06	203.85	1038.27
14	63	29	-14.28	-35.22	502.90	203.85	1240.60
15	63	43	-14.28	-21.22	303.01	203.85	450.38
16	50	64	-27.28	-0.22	6.06	744.08	0.05
17	50	79	-27.28	14.78	-403.10	744.08	218.38
18	38	79	-39.28	14.78	-580.44	1542.74	218.38
SUM	1391	1156			3216.89	6589.61	9181.11
MEAN	77.28	64.22					
				VARIANCE		366.09	510.06
				STD. DEV.		19.13	22.58
				PPMC	0.41		

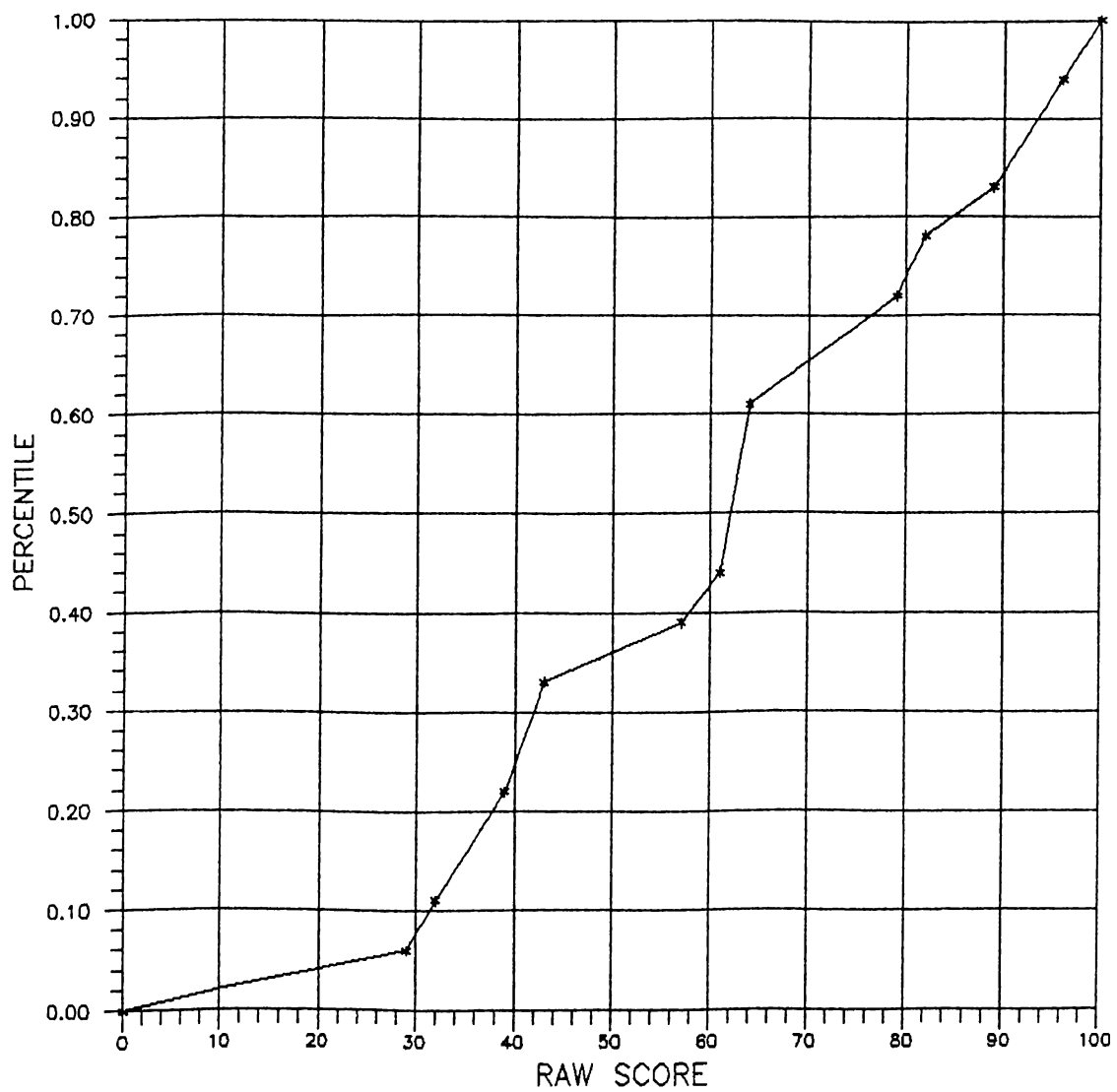
## APPENDIX H.1

PERCENTILE & RAW SCORES FOR MULTIPLE-CHOICE TEST  
Pass mark for multiple choice= 72



## APPENDIX H.2

PERCENTILE & RAW SCORES FOR RETELLING  
Pass mark for retelling = 63





APPENDIX H.3

PERCENTILE & RAW SCORES FOR BUSEL READING EXAM.  
Pass mark for Busel Reading Exam = 71

