

**FARKLI TEST YÖNTEMİ İNGİLİZCE OKUMA BECERİSİNİN ÖLÇÜLMESİNDE  
FARKLI SONUÇLAR YARATIR MI?**

**DO DIFFERENT TESTING METHODS  
YIELD DIFFERENT SCORES  
IN ASSESSING READING  
COMPREHENSION?**

**EBRU CİHANGİRLİ  
(Yüksek Lisans Tezi)**

**Eskişehir, 2000**

DO DIFFERENT TESTING METHODS YIELD DIFFERENT SCORES IN  
ASSESSING READING COMPREHENSION?

FARKLI TEST YÖNTEMİ İNGİLİZCE OKUMA BECERİSİNİN ÖLÇÜLMESİNDE  
FARKLI SONUÇLAR YARATIR MI?

Ebru Cihangirli

THESIS OF MASTER OF ARTS  
Advisor: Prof. Dr. Zülal Balpınar

Eskişehir  
Anadolu Üniversitesi Sosyal Bilimler Enstitüsü  
Şubat 2000

## YÜKSEK LİSANS TEZ ÖZÜ

FARKLI TEST YÖNTEMİ İNGİLİZCE OKUMA BECERİSİNİN ÖLÇÜLMESİNDE  
FARKLI SONUÇLAR YARATIR MI?

Ebru Cihangirli

İngiliz Dili Eğitimi Anabilim Dalı  
Anadolu Üniversitesi Sosyal Bilimler Enstitüsü  
Danışman: Prof. Dr. Zülal Balpınar

Bu çalışmanın başlıca amacı, İngilizce okuma becerisinin değerlendirilmesinde test yönteminin alınacak nota etki eden bir faktör olup olmadığını araştırmaktır. Bunu araştırmak için, aynı içeriği ölçen “Açık Uçlu, Çoktan Seçmeli, ve Doğru-Yanlış” tarzda testler hazırlanmıştır. Bu testler aynı bilgi seviyesindeki 3 öğrenci grubuna verilerek elde edilen veriler hem kendi grupları içinde, hem de diğer gruplarla karşılaştırılmıştır.

Bu çalışmada veriler 1999-2000 öğretim yılı Güz döneminde Osmangazi Üniversitesi Hazırlık Programı’na devam etmekte olan 60 orta seviye öğrenciden alınmıştır. Öğrenciler “A, B, C” olarak 3 gruba ayrılmış, Grup A sınavları “Açık Uçlu”, Grup B “Çoktan Seçmeli”, Grup C “Doğru-Yanlış” metodla almıştır. Çalışmanın başında öğrencileri gruplandırmak için verilen Seviye Tespit Sınavı, çalışmanın sonunda da verilerek dil seviyelerindeki ilerleme tespit edilmiştir.

Çalışmada kullanılan metinler ve bunlara dayalı olan “Çoktan Seçmeli” sorular “Developing Reading Skills. Intermediate” (Markstein, 1981), adlı kitaptan alınmış, “Açık Uçlu” ve “Doğru-Yanlış” metoduyla verilen sınavlar, “Çoktan Seçmeli” soruların “Açık Uçlu” ve “Doğru-Yanlış”a dönüştürülmesiyle elde edilmiştir. Sınavlardan elde edilen veriler “t-test” yöntemiyle karşılaştırılmıştır.

Testlerden elde edilen sonuçlara göre; test metodunun dil seviyesindeki ilerlemeye paralel olarak skora etki eden bir faktör olmaktan çıktığı saptanmıştır. Sonuçlardaki farklılığın dil seviyesinin düşük olduğu dönemlerde görüldüğü, hedef dilde ilerleme kaydedildikçe bu farklılığın azaldığı ve zamanla ortadan kalktığı sonucuna ulaşılmıştır.

## THESIS OF MASTER OF ARTS

## ABSTRACT

DO DIFFERENT TESTING METHODS YIELD DIFFERENT SCORES IN  
ASSESSING READING COMPREHENSION?

Ebru Cihangirli

English Language Education Department

Anadolu University Social Sciences Institute

Advisor: Prof. Zulal Balpınar

The main aim of this study is to search whether testing method has a significant effect on testing reading comprehension. In order to do this three open-ended, multiple-choice, and true-false reading tests, which are based on the same passage and which test the same content, were prepared. These tests were given to three intermediate level student groups. The data obtained from the tests was compared within and between the groups.

In this study the data was collected from sixty intermediate level students attending Intensive English Program of Osmangazi University, Eskişehir, Turkey. These students were divided into three groups: "Group A", "Group B", "Group C". Group A took the tests with "Open-ended Test Method"; Group B took the tests with "Multiple-Choice Test Method"; Group C took the tests with "True-False Test Method". The placement test, which was given at the beginning of the study, was also given at the end of the study to observe the improvement in students' linguistic levels.

The passages and the multiple-choice questions, used in this study, were taken from "Developing Reading Skills. Int." (Markstein, 1981). The Open-ended and True-false tests were prepared by converting Multiple-choice questions to Open-ended and True-false questions. The data obtained from the study was correlated with t-test.

According to the results of the tests, it was found that, by the time students became more proficient, the test method became an insignificant factor. When the proficiency level of the students low, different scores were obtained from different testing methods. When the students became more proficient, it was observed that the difference became insignificant.

## JÜRİ VE ENSTİTÜ ONAYI

Ebru Cihangirli'nin "Do Different Testing Methods Yield Different Scores in Assessing Reading Comprehension?" başlıklı tezi ...21-4-2000.. tarihinde, aşağıdaki jüri tarafından Lisansüstü Eğitim Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca, İngiliz Dili Eğitimi Anabilim Dalı'nda Yüksek lisans Tezi olarak değerlendirilerek kabul edilmiştir.

İmza

Tez Danışmanı: Prof.Dr. Zülal Balpınar

Üye : Doç.Dr. İlknur Kesik

Üye : Doç.Dr. Ferhan Odabaşı

Prof .Dr. Enver Özkalp

Enstitü Müdürü

## ACKNOWLEDGEMENTS

I would like to express my gratitude to Prof. Zülal Balpınar, my thesis advisor, for her support and invaluable suggestions for the preparation of this thesis.

I am also grateful to my colleague, my best friend, Güzde Uludağ for her utmost care, patience and support in every process of this thesis.

I must express my special thanks to Harun Sönmez for his help with the statistical data analysis of this thesis.

My sincerest thanks also go to my family, several of my colleagues and my friends who encouraged me and gave me too much morale support in every stage of this thesis. It was on your wings that I learned to fly!

Finally, I wish to thank to my students who were involved in this study.

## TABLE OF CONTENTS

### SECTIONS

#### Pages

Özet	iii
Abstract	iv
Değerlendirme Kurulu ve Enstitü Onayı	v
Acknowledgements	vi
Özgeçmiş	vii
Table of Contents	viii
List of Tables	x

### CHAPTER I

#### INTRODUCTION

1.1. Background to the problem .....	1
1.1.1. Open-ended Tests .....	3.
1.1.2. Multiple-choice Tests .....	4.
1.1.3. True-false Tests .....	6.
1.2. Aim and scope of the study .....	7.
1.3. Limitations of the study .....	8.

### CHAPTER II

#### LITERATURE REVIEW

2.1. Reading comprehension .....	10
2.2 Related studies Conducted .....	11.

### CHAPTER III

#### METHODOLOGY

##### Introduction

3.1. Subjects .....	15
3.1.1. Group A .....	16.
3.1.2. Group B .....	16.
3.1.3. Group C .....	16.
3.2. Instruments .....	17.
3.2.1. Michigan Placement Test .....	17.
3.2.2. Tests .....	17.
3.3. Data Collection Procedure .....	18.
3.4. Data Analysis .....	21.

**CHAPTER IV  
DATA ANALYSIS**

**Introduction**

4.1. Statistical Results .....	22
4.1.1. Tests .....	22
4.1.2. Comparison of the tests within the groups .....	22
4.1.3. Comparison of the tests between the groups .....	27
4.1.4. T-test for Dependent Samples .....	32

**CHAPTER V  
DISCUSSIONS AND CONCLUSIONS**

5.1. Summary .....	36
5.2. Discussion .....	38
5.3. Suggestions for further studies .....	41

**APPENDICES**

Appendix A1 >>>>> First Test's Passage .....	43
Appendix A2 >>>>> Open ended questions for the First Test .....	44
Appendix A3 >>>>> Multiple-choice questions for the First Test .....	45
Appendix A4 >>>>> True-false questions for the First Test .....	47
Appendix B1 >>>>> Second Test's Passage .....	48
Appendix B2 >>>>> Open ended questions for the Second Test .....	50
Appendix B3 >>>>> Multiple-choice questions for the Second Test .....	51
Appendix B4 >>>>> True-false questions for the Second Test .....	53
Appendix C1 >>>>> Third Test's Passage .....	54
Appendix C2 >>>>> Open ended questions for the Third Test .....	56
Appendix C3 >>>>> Multiple-choice questions for the Third Test .....	57
Appendix C4 >>>>> True-false questions for the third Test .....	59
Appendix D >>>>> Groups' first, second, third test and Pre&post placement test scores .....	60

<b>BIBLIOGRAPHY .....</b>	<b>61</b>
---------------------------	-----------



## LIST OF TABLES

TABLE 4.1. Group A (Test 1 vs Test 2) T-test for dependent samples .....	23
TABLE 4.2. Group A (Test 1 vs Test 3) T-test for dependent samples .....	23
TABLE 4.3. Group A (Test 2 vs Test 3) T-test for dependent samples .....	24
TABLE 4.4. Group B (Test 1 vs Test 2) T-test for dependent samples .....	24
TABLE 4.5. Group B (Test 1 vs Test 3) T-test for dependent samples .....	25
TABLE 4.6. Group B (Test 2 vs Test 3) T-test for dependent samples .....	25
TABLE 4.7. Group C (Test 1 vs Test 2) T-test for dependent samples .....	26
TABLE 4.8. Group C (Test 1 vs Test 3) T-test for dependent samples .....	26
TABLE 4.9. Group C (Test 2 vs Test 3) T-test for dependent samples .....	27
TABLE 4.10. First Test (Group A vs Group B) T-test for independent samples .....	27
TABLE 4.11. First Test (Group A vs Group C) T-test for independent samples .....	28
TABLE 4.12. First Test (Group B vs Group C) T-test for independent samples .....	28
TABLE 4.13. Second Test (Group A vs Group B) T-test for independent samples .....	29
TABLE 4.14. Second Test (Group A vs Group C) T-test for independent samples .....	29
TABLE 4.15. Second Test (Group B vs Group C) T-test for independent samples .....	30
TABLE 4.16. Third Test (Group A vs Group B) T-test for independent samples .....	30
TABLE 4.17. Third Test (Group A vs Group C) T-test for independent samples .....	31
TABLE 4.18. Third Test (Group B vs Group C) T-test for independent samples .....	31
TABLE 4.19. Pre-Placement Test vs Post Placement Test (Group A) T-test for dependent samples .....	32
TABLE 4.20. Pre-Placement Test vs Post Placement Test (Group B) T-test for dependent samples .....	33

TABLE 4.21. Pre-Placement Test vs Post Placement Test (Group C)	
T-test for dependent samples .....	33
TABLE 4.22. Interrelationship of the scorers (Open-ended Test 1)	
T-test for dependent samples .....	34
TABLE 4.23. Interrelationship of the scorers (Open-ended Test 2)	
T-test for dependent samples .....	34
TABLE 4.24. Interrelationship of the scorers (Open-ended Test 3)	
T-test for dependent samples .....	35
TABLE 5.1. The overall mean scores of the tests .....	37

# 1. INTRODUCTION

## 1.1. Background to the problem

Assessment of student learning requires a number of techniques for measuring student achievement. But assessment, as Aksu (1997) describes, is a systematic process that begins with the identification of learning goals and ends with a judgement concerning the extent to which those goals have been achieved.

Testing is one of the ways of getting information about students' performance on language skills. Testing has many advantages, as Basanta (1995) states:

- \*tests show teachers how successful they have taught

- \*tests provide washback for teachers to adjust, change their teaching style or course content.

- \*tests inform students about their progress in language

- \*tests show teachers in which areas of their teaching needs revision and recovery

In summary, the main goal of classroom testing is to obtain valid, reliable and useful information concerning student achievement.

Madsen (1983) states testing is an important part of every teaching and learning experience. According to Shohamy (1993), in the process of testing two major factors interact: one is the trait, the other is the method. Trait refers to the knowledge that is being measured, such as writing, grammar, reading comprehension. Method refers to the procedure or technique that is used to assess the trait. Since there are many methods and procedures to assess a given trait, these methods may have a differing effect on the trait being measured and, consequently, on the scores that test takers obtain as a result of the test taker's familiarity with a specific testing method or of any other effect that the

method may have on the trait. Thus, one characteristic of a good test is that the method has little effect on the trait. A bad test, on the other hand, is when the method has a strong effect on the trait being measured, and on the test taker's scores on such tests.

Baker (1989) states language testing is a complicated subject and much of this complication stems from problems of description and measurement which are particularly acute in linguistic and psychological investigation.

Madsen (1983) states that language testing is seen in two general categories: a) tests of language subskills and, b) tests of communication skills. Tests of language subskills measure the separate components of English, such as vocabulary, grammar and pronunciation. Tests of communication skills show how well students can use the language in exchanging ideas and information such as reading, speaking, writing and listening.

Being one of the communication skills, reading plays an important role in foreign language teaching and testing. Reading is known as a receptive skill. Unlike speaking and writing, reading comprehension is an unobserved skill; so the tester does not have an actual language sample to analyse and work with and does not know what is really happening in the mind of the test taker while he/she is reading (Shohamy, 1983). When people write and speak, we see and hear; when they read and listen there is often nothing to observe (Hughes, 1989). Thus, reading comprehension can only be evaluated directly when the tester imposes testing methods such as Multiple-Choice, Open-Ended, True-False, etc. which he/she believes will tap the reading comprehension trait. An important issue in testing unobserved skills, such as reading comprehension, is how the trait can be assessed with minimal effect caused by the testing method (Shohamy, 1993).

Gordon & Hanauer (1995 ) also points out the difficulty of measuring mental model construction. He thinks that it is problematic because mental processes are unobservable.

Since testing reading is about measuring the mental process of reader, it has always been problematic and full of questions. Testing reading comprehension methods vary, as the number of methods vary, it gets more difficult to decide on the method. Bachman (1990) states that performance on language tests varies as a function both of an individual's language ability and of the characteristics of the test methods.

Shohamy (1984) says research on testing methods is especially important for testing reading comprehension, since the mental processing of that trait is not known nor understood. Reading comprehension process takes place in the brain which is impossible to see. It can be analysed indirectly via testing methods such as Multiple-Choice, Open-Ended, True-False, Cloze, etc.

In this study only Open-ended, Multiple-Choice and, True-False test types will be analysed:

#### **1.1.1. Open-ended Tests**

The term Open-ended is used to refer to those questions which elicit a completely subjective response on the part of the testees. The response required may range from a one word answer to one or two sentences (Heaton,1988).

Open-ended type questions require the candidate to write down answers, and allow the candidate some freedom of expression. Answers are not provided for the students as in Multiple-choice tests; therefore, production can be seen easily (Weir,1990).

Although there are no studies directly examining how Open-ended questions affect readers' ability to demonstrate comprehension, it is highly likely that Open-ended tests could receive some of the same criticisms as Multiple-choice tests. Open-ended questions may encourage bottom up processing of a text in that test takers may derive meaning solely from textual features. Also, Open-ended questions that are used to check reading comprehension in many textbooks often test only isolated facts and details. They do not require that readers integrate information from different levels of the passage (Wolf, 1993).

In this study, with Open-Ended questions it is referred to ones which require one word answer or short sentences.

### **1.1.2 Multiple-choice Tests**

One of the best ways to test reading comprehension is a reading passage followed by Multiple-Choice questions (Madsen,1983). Weir (1990) states that a multiple-choice item is usually established in such a way that the candidate is required to select the answer from a number of given options, of which only one is correct. Because agreement has already been made regarding the correct answer for each item, the marking process is totally objective. Multiple-Choice questions are used to assess the subject's comprehension of passages because they are familiar to subjects and easy to score (Wolf,1993). Multiple-Choice items are favoured in many cases because their scoring can be reliable, rapid, and economical (Cohen,1994). Killoran (1992) also states that Multiple-Choice type questions are an effective assessment device when used correctly.

Although Multiple-Choice items are favoured, their advantages and disadvantages should be evaluated:

**Advantages:** In Multiple Choice tests, there is almost complete marker reliability. Candidates' marks can not be affected by personal judgement. As well as being reliable, marking is simple, more rapid and often more cost effective than other forms of written tests (Weir,1990), (Coller,1979), (B.J.Carroll, in Mason 1984). Because items can be pre-tested easily, it is always possible to estimate in advance the difficulty level of each item and that of the test as a whole. By Multiple-Choice tests, the candidates know what is required of them. (Weir;1990), (Coller,1979), (Mason,1984), (Rivers,1978).

**Disadvantages:** If a candidate gets a Multiple-Choice item wrong because of some flaw in the question, the answer sheet on which he records his answers will not reveal this fact. In addition, it is not known whether a candidate's failure is due to lack of comprehension of the text or lack of comprehension of the question (Weir,1990).

As Johns, Juinman, Perkins, and Sones point out, any of the test items can be answered quickly without reading the passage; that is, they are not passage dependent. Another criticism of multiple choice tests is that they do not require careful reading of the passages to answer the questions ; and Swaffar and Waltermann point out, test takers can rely on the recognition of a few key words in the passage or on clues from other questions to perform better than they would merely by chance. (in Wolf,1983).

As Conor and Read point out that test takers rely on clues from the other questions, or syntactic or semantic relationship between the stem and the choices, thus

scoring higher than they would merely by chance (in Wolf,1993). Likewise, Swaffar and Waltermann criticise Multiple-Choice tests because they encourage a bottom up strategy for reading. Thus test takers can often perform well on Multiple-Choice tests without having truly understood the passage (in Wolf, 1993).

There is considerable doubt about their validity as measures of language ability. Answering Multiple-Choice items is an unreal task, as in real life one is rarely presented with four alternatives (Weir,1990). Also Multiple-Choice tests take much longer and are more expensive and difficult to prepare (Weir,1990).

### **1.1.3. True- False Tests**

The True-false test is widely used in testing reading comprehension. The scoring of these tests is straightforward and quick, thus, the scores obtained are very reliable. True-false tests are of considerable use for inclusion in class progress tests chiefly because, unlike M.C test items, they can be constructed easily and quickly (Heaton,1988).

However, Heaton (1988) states two main disadvantages: firstly, it can encourage guessing, since testees have a 50 percent chance of giving a correct answer for each item. Secondly, as the base score is 50 percent and thus the average test difficulty generally in the region of 75 percent, the test may fail to discriminate widely enough among the testees unless there are a lot of items. By adding a third question in addition to the True-False option, e.g.: "True, False, Not Stated", and with some penalties, e.g.: deducing one mark from the score and awarding 2 marks for each correct answer, the guessing problem can be solved.



## 1.2. Aim and Scope of the Study

In this study, the effect of test method on reading comprehension test scores will be examined. The aim of this study is to find out whether different testing methods yield different scores in the assessment of reading comprehension, ( and whether there is any difference among three reading comprehension methods: Open-Ended, Multiple-Choice, True-False).

By investigating the effect of test type in reading via several tests, it is aimed to contribute to testing reading research. The aim is to reveal whether there is a difference in the scores of three intermediate level group students, who take the same text with different testing methods. Therefore, in this study the main research questions will be as follows:

1. Is there a significant difference between the mean scores of:

\*Group A's (Students who take the tests with Open-Ended Method) 1<sup>st</sup> and 2<sup>nd</sup> tests?

\*Group A's 1<sup>st</sup> and 3<sup>rd</sup> tests?

\*Group A's 2<sup>nd</sup> and 3<sup>rd</sup> tests?

2. Is there a significant difference between the mean scores of:

\*Group B's (Students who take the tests with Multiple-Choice method) 1<sup>st</sup> and 2<sup>nd</sup> tests?

\*Group B's 1<sup>st</sup> and 3<sup>rd</sup> tests?

\*Group B's 2<sup>nd</sup> and 3<sup>rd</sup> tests?

3. Is there a significant difference between the mean scores of:

\*Group C's (Students who take the tests with True-False method) 1<sup>st</sup> and 2<sup>nd</sup> tests?

- \*Group C's 1<sup>st</sup> and 3<sup>rd</sup> tests?
  - \*Group C's 2<sup>nd</sup> and 3<sup>rd</sup> tests?
4. Is there a significant difference between the mean scores of:
- \*Group A's 1<sup>st</sup> and Group B's 1<sup>st</sup> tests?
  - \*Group A's 1<sup>st</sup> and Group C's 1<sup>st</sup> tests?
  - \*Group B's 1<sup>st</sup> and Group C's 1<sup>st</sup> tests?
5. Is there a significant difference between the mean scores of:
- \*Group A's 2<sup>nd</sup> and Group B's 2<sup>nd</sup> tests?
  - \*Group A's 2<sup>nd</sup> and Group C's 2<sup>nd</sup> tests?
  - \*Group B's 1<sup>st</sup> and Group C's 1<sup>st</sup> tests?
6. Is there a significant difference between the mean scores of:
- \*Group A's 3<sup>rd</sup> and Group B's 3<sup>rd</sup> tests?
  - \*Group A's 3<sup>rd</sup> and Group C's 3<sup>rd</sup> tests?
  - \*Group B's 3<sup>rd</sup> and Group C's 3<sup>rd</sup> tests?
7. Is there a significant difference between the mean scores of:
- \*Group A's pre-placement and post-placement tests?
  - \*Group B's pre-placement and post-placement tests?
  - \*Group C's pre-placement and post-placement tests?
8. Is there an effect of proficiency on test type effect

### 1.3. Limitations of the study

This study was conducted with a limited number of subjects attending Osmangazi University Intensive English Program. In this study 3 testing reading comprehension methods were used: Open-Ended, Multiple-Choice and True-False;

other testing methods were not included in this study. This study was limited to intermediate level Turkish EFL students and to the reading skill.

In the following chapter, related studies conducted will be presented.

## **2. LITERATURE REVIEW**

### **2.1. Reading Comprehension**

In foreign language learning and teaching reading comprehension is considered to be one of the most important skills. Rivers (1968:259) states that, reading is a most important activity in any language class, not only as a source of information and a pleasurable activity, but also as a means of consolidating and extending one's knowledge of the language. Bernhardt (in Wolf, 1993:116) defines reading comprehension as a "constructive" and "active process" that entails "relating new and incoming information to information already stored in memory". Rivers (1968:259) defines reading as an important activity for expanding knowledge of a language.

Wolf (1993:473) states that, reading has a dual role in the foreign language acquisition process as a source of information and a source of input. Due to the important role of reading learners must be able to comprehend written texts.

As well as the importance of teaching of reading, testing reading comprehension should also be given much importance. Dollerup et al (1982, in Nevo 1989:200) indicate a relationship between problem solving strategies in reading comprehension tests and the reading process itself. Two elements are seen as being involved in the reading process: reading ability and linguistic competence. Reading ability refers to both native and the target language. It is assumed that if a reader is good in his first language, he will be good in the target language reading, as well. By linguistic competence it is referred to coping with reading problems.

Assessing learners' comprehension is problematic in that the processes by which readers create meaning from printed texts are invisible and therefore can not be measured directly ( Wolf, 1993).

## **2.2. Related Studies Conducted**

Research on testing methods is especially important for testing reading comprehension since the mental processing of that trait is not known and understood. Reading comprehension in a second language is even more of a puzzle because it involves unknown aspects from first and second language. Reading comprehension can only be evaluated indirectly. Traits, which are evaluated indirectly, put a heavy burden on the testing method and, therefore, may create greater variations in the scores obtained as a result of the different methods. An important issue in testing unobserved skills, such as reading or listening comprehension, is how the trait can be assessed with minimal effect caused by the testing method (Shohamy,1984:147-149).

A great deal of research has been carried out to find out if different testing methods affect the trait being measured and test takers' scores. These researches can be categorised under two debates: ones about testing reading comprehension with Multiple-choice and Open-ended methods and, ones about the other testing methods and other language skills.

Madsen (1983) points out that testing reading comprehension is one of the most integrative and challenging kinds of reading test types. Johnston (in Davey, 1989) agrees with Madson and says that the assessment of reading comprehension has been particularly problematic due to complex interactions among readers, texts, and tasks during comprehension processing. As Fair and Coney (1986) cited, reading processes

are not directly observable and must be inferred from samples of behaviours assumed to reflect real reading (in Davey, 1989).

The effect of test format on test performance (method effect) has been examined quite extensively. The research indicates that the test method itself has a significant effect on test scores (Birenbaum&Tatsuoka 1987; Gordon, 1987; Samson, 1983; Shohamy, 1984; Ward, Fredricksen, Carlson, 1980 in Gordon&Hanauer,1995). Campbell and Fiske (1959, in Gordon&Hanauer,1995)) note if the method effect is strong enough, the information obtained from the test will be invalid.

Several studies have been carried out to find the effect of testing methods on the test scores. Kendall (et all 1980), used four different tasks (Multiple-choice, recall, and maze tasks) to measure reading comprehension using the same text, data analyses revealed significant effect for the methods. Students scored higher on Multiple-choice and Maze tests..

Stathmann (1979) also tried to find out the significance of test type. In a study he carried out, he prepared a test, which he included, both Multiple-choice and Short-answer questions. Half of the questions were Multiple-choice and the other half was Short-answer. In his study he found out that seventy-five percent of the students got the same proportion of Multiple-choice and Short-answer questions correct. Twenty percent did better on the Short-answer questions and five percent did better on Multiple-choice questions.

Bormuth (1967, in Bensoussan 1984) has shown that it is possible to arrive at equivalent cloze and multiple-choice comprehension test percentage scores using the same texts. He found out that both test methods measured the same skill.

Another study by Samson (1983 in Shohomy, 1984) investigated three methods for testing reading comprehension: Multiple-choice , Open-ended and summary writing. As a result of the study, there were no major differences, which means that the same trait was measured by all the three methods. Multiple-choice was found to be the easiest method for the test takers, followed by the Open-Ended questions and then by the summary procedure.

Shohomy (1984) also examined the effect of the Multiple-choice method and Short-answer test method on reading comprehension. The results indicated that each of the testing methods produced different degrees of difficulty for the test taker and the method had a significant effect on student scores in reading comprehension.

In another investigation of Multiple-choice and Open-Ended methods on EFL reading tests, Gordon (1987 in Gordon&Hanauer,1995) found that questions themselves provide clues to understanding the text. This means that not only does the test format affect performance on reading tests but, in fact, the test task has an effect on the meaning the test taker has constructed of the stimulus text as well.

Statman (1988) compared different methods of MC testing and showed that students scored higher on Multiple-choice items which have the format of a question with one of four distracters giving the correct answer, than on the common format in which the testee has to complete a sentence stem by choosing one of four distracters.

Davey (1989) states that results from several studies indicated that Multiple-choice question scores were generally higher than Open-Ended question scores for all readers. (Kendall et all,1980; Samson 1983, in Shohamy 1984; Davey 1987, in Davey 1989). Results from several studies, conducted by Davey (1987,1988) and Mac Ready (1986), state that M.C items appear to be assessing comparable processes and skills for

both good and poor readers, and this may not be the case for free response items (in Davey, 1989).

Alderson (1983) compared different methods of cloze testing and found that changing the deletion frequency produced a test which measured different abilities, unpredictably. He also found that changing the text resulted in a different measure of English as a foreign language proficiency, such that a more difficult text provided a better measure of core proficiency. Changes in scoring procedures also resulted in different validities of the cloze test.

Bachman (1983 in Shohamy, 1984:148) administered cloze tests of fixed-ratio deletion and rational deletions and found differences in difficulty level and factor structures in the two procedures.

Klein-Braley (1981, in Gordon&Hanauer) claims that it makes no sense to assume perfect equivalence of cloze tests which are based on different passages since she found that the deletion rules used in cloze tests made a difference in the proficiency scores of EFL advanced students.

Porter (1978) compared different methods of cloze testing using the same passage, and data analyses revealed that each form yielded different results. In other words, the score of the test takers on the same test is likely to change depending on the form and found that the five possible cloze forms constructed over the same passage do not yield equivalent results. In other words, an individual's score on the same test is likely to change depending on which of the five forms have been administered.

With this study, it is aimed to learn more about the process of testing reading comprehension, which can be observed indirectly by investigating the interrelationship between different testing methods and scores.



### **3. METHODOLOGY**

#### **Introduction:**

In this chapter, the design of the study which including subjects, the instruments used, and data collection procedures, will be discussed.

#### **3.1 Subjects**

Sixty-five intermediate level students attending Osmangazi University Intensive English Program, Eskisehir were chosen as the subjects of this study. As the researcher herself has been instructing one of the intermediate level classes, this level was chosen. Students who enter Osmangazi University Intensive English Program, take a placement test and from the scores they receive, they are placed into elementary, pre-intermediate, intermediate, or upper-intermediate class. The Michigan Placement test was given to all groups to have better groups and regroup. Sixty-five of the students were chosen as the subjects of this study, then they were divided into three distinct research groups: Group A, students who took three Open-Ended reading comprehension tests; Group B, students who took three Multiple-Choice reading comprehension tests; and Group C, students who took three True-False reading comprehension tests. Subjects' pre-placement test scores ranged from 35 to 45. Their native language is Turkish and their ages range between 17 and 19.

### **3.1.1. Group A**

Twenty students attending preparatory class D were chosen as Group A. The students of this class were given three open ended type intermediate level reading comprehension tests. During the study, two students who missed one of the quizzes, were excluded from the analysis and the final sample consisted of 20 subjects. Group A was chosen randomly.

### **3.1.2. Group B**

Twenty students attending preparatory class G were chosen as Group B. The students of this class were given three Multiple Choice type intermediate level reading comprehension tests. Group B was chosen randomly.

### **3.1.3. Group C**

Twenty-three students attending preparatory class E, were chosen as Group C. The students of this class were given three True-False type intermediate level reading comprehension tests. During the study, three students who were absent from one of the tests, were excluded from the analysis and the final sample consisted of 20 subjects. Group C was chosen randomly.

The study was conducted at Osmangazi University Intensive English Program, in the students' usual classrooms during their usual class hours to lower stress and anxiety. All of the groups took the same texts with different testing techniques.

## **3.2. Instruments**

In this study, which aims to determine the significance of test method on reading comprehension test scores, the following instruments were used:

### **3.2.1. Michigan Placement Test**

This is a standard placement test which is composed of three parts; listening comprehension, grammar, vocabulary and reading comprehension parts. There are a total of 100 questions and scores are calculated on a 100 point scale.

### **3.2.2. Tests**

A total of three tests were given to students. They were graded out of 100. All of the texts were taken from an intermediate level reading text book.

(Markstein, 1981) The length of each text was about 280 words long respectively. The original passages included 10 Multiple Choice questions. Open-Ended and True-False type questions were prepared by the researcher. In the light of the findings obtained, the final form of the tests were prepared. In order to have better tests and questions, two colleagues read the texts and the questions. To have linguistic and content validity, three test methods (Open-Ended, Multiple-Choice, True-False), based on the same passage, were piloted on three groups of upper intermediate level students.

The reliability coefficients of each test was computed separately through Split-Half formula and the results were found to be as follows:

First Open Ended test  $r: 0,71$  ; Second Open Ended test  $r: 0,84$  ; Third Open Ended test  $r: 0,82$ .

First Multiple Choice test  $r: 0,69$  ; Second Multiple Choice test  $r: 0,69$  ; Third Multiple Choice test  $r: 0,90$ .

First True False test r: 0,66 ; Second True False test r: 0,82 ; Third True False test r: 0,82 .

### **3.3. Data Collection Procedure**

The procedure lasted for one semester. It started at the beginning of the Fall Semester of 1999-2000 academic year. The procedure consisted of the following stages:

a) The administration of pre-placement test, b) the selection of reading passages which will be used in the study, c) The construction of the Open-Ended and True-False versions of the tests, d) the pilot study, e) the administration of reading tests and the scoring procedures, and f) the administration of the post placement test.

#### **a) The administration of pre-placement test:**

The Michigan Placement Test was given at the beginning of the study to determine intermediate level students. The scores of the students ranged from 35 to 45.

#### **b) The selection of the reading passages:**

Three reading passages were selected after careful consideration of a number of reading passages. The reading passages used in the study were taken from "Developing Reading Skills-Intermediate" by Markstein (1981). The criteria used by Testing Unit of Osmangazi University Preparatory School was taken into consideration during the selection process.

### **c) Construction of the tests and the pilot study:**

After the selection of the reading passages, two teachers teaching intermediate level students at Osmangazi University Intensive English Program examined and found them suitable in terms of content and level. After that, Open-Ended, and True-False questions were constructed for each test. Questions were constructed by converting each Multiple-Choice question to Open-Ended, and True-False forms. The tests differed in testing method only.

### **d) The Pilot Study:**

In order to check content and linguistic validity of the tests, they were piloted in three upper intermediate level groups of Osmangazi University Prep School. The data obtained from the pilot study was taken into consideration by two experienced prep school teachers and the researcher. Necessary changes were made, and some items were revised.

### **e) The test administration and scoring procedure:**

The final form of the three different versions of the first test was given to three intermediate level groups, on the same day. The first test was given on the 3<sup>rd</sup> week of the semester. The second test was given under the same conditions just after the 1<sup>st</sup> midterm, which was the 6<sup>th</sup> week of the semester. The third test was given on the 9<sup>th</sup> week of the semester. Multiple Choice and True-False groups were given ten minutes with the suggestion allowing one minute of testing time per item ( Harris, 1969). Considering the time the students spend for writing the answers of the questions. Group A (Students who took the tests with Open-Ended method) was given 15 minutes. The class teachers conducted the tests. All of the groups read the same text but had different types of tests. During the study, Group A took the tests with Open-Ended

method, Group B with Multiple-Choice method and Group C with True-False method.

One week after the administration of the 3<sup>rd</sup> quiz, the placement test, which was given at the beginning of the study, was given again as post placement-test to determine the proficiency improvement within the groups.

Multiple-Choice and True-False reading comprehension tests were scored by the researcher. In order to have inter-rater reliability, that is, “the degree of correspondence and consistency between the sets of scores awarded by different judges using a given technique” (Enginarlar, 1991), Open-Ended reading comprehension tests were scored by the researcher and another prep school teacher of Osmangazi University Intensive English Program.

For True-false and Multiple-choice reading comprehension tests, there was only one answer given, there was no half mark. When marking Open-ended items, as Heaton (1988) suggested, two marks for each correct answer were awarded. Thus, Open-Ended comprehension tests, for a correct response 10 points was granted and for a partially correct response 5 points was granted. Spelling and grammar mistakes (as long as the meaning was not obstructed) were ignored since the aim of the test was only to assess comprehension ability.

#### **f)The administration of the post-placement test:**

The Michigan Placement Test was given at the end of the study as the post placement test to determine the improvement in students' English proficiency. The scores of the students ranged from 55 to 65 .

### **3.4 Data Analysis**

The purpose of this study is to find out whether different testing methods yield a difference in assessing reading comprehension. In order to find out the answer, three intermediate level reading comprehension passages and three different testing methods were chosen: Open-ended, Multiple-choice and True-false tests were constructed for each passage. Subjects were categorised as Group A, Group B and Group C. Group A took the tests in Open-ended format, Group B took the tests in Multiple-choice format and Group C took the tests in True-false format. The same procedure was repeated for three passages.

A placement test was given at the beginning and at the end of the study. The aim of the first placement test was to regroup students at Osmangazi University English Program. The aim of the second placement test was to see the improvement in their scores.

To find out whether there is any difference between the scores of 3 intermediate level student groups' test scores, who were given the same passages with different testing methods, the results were correlated within and between the groups. The reading comprehension tests were graded out of 100, the averages for each quiz for each group were calculated, and t-test was applied to see if there were significant differences between the groups' scores.

## 4. DATA ANALYSIS

### Introduction

In this chapter, the results obtained from various instruments were analysed through t-tests; t-test for dependent samples, t-test for independent samples, and t-test for the differences in the scores between pre and post placement tests at the 0.05 level of significance. First of all the mean scores of each group's reading comprehension tests were compared within groups for each test. Then, the test scores were compared between the groups for each test. Finally, pre and post placement test scores of students were compared within the groups.

### 4.1. Statistical Results

#### 4.1.1. Tests

During the study, three tests were administered to groups. The purpose of these tests was to evaluate the reading progress and determine the differences within and among the groups. Three same passages were given to groups in three testing methods.

#### 4.1.2. Comparison of the tests within groups:

The following hypothesis will try to answer Research

#### Question 1

##### Group A

**Hypothesis 0:** There is no significant difference between the mean scores of group A's first and second tests.

**Hypothesis 1:** There is a significant difference between the mean scores of group A's first and second tests.

In order to test the null hypothesis, a t-test was run and the results showed that there was a significant difference ( $t = -5,00450$   $p = ,000079$ ), since  $p < 0.05$ .



**Table 4.1**  
**Group A (Test 1vs Test 2)**  
T-test for dependent samples

Variable	Mean	Std.Dv.	N	t	df	p
Group A 1	53,25000	9,215976	20	-5,00450	19	,000079
Group A 2	66,50000	9,880869				

At 0,05 level of significance

**Hypothesis 0:** There is no significant difference between the mean scores of group A's first and third tests.

**Hypothesis 1:** There is a significant difference between the mean scores of group A's first and third tests.

In order to test the null hypothesis, a t-test was run and the results showed that there was a significant difference ( $t = -12,3784$   $p = ,000000$ ), since  $p < 0.05$ .

**Table 4.2**  
**Group A (Test 1vs Test 3)**  
T-test for dependent samples

Variable	Mean	Std.Dv.	N	t	df	p
Group A 1	53,25000	9,215976	20	-12,3784	19	,000000
Group A 3	78,25000	8,155560				

At 0,05 level of significance

**Hypothesis 0:** There is no significant difference between the mean scores of group A's second and third tests.

**Hypothesis 1:** There is a significant difference between the mean scores of group A's second and third tests.

In order to test the null hypothesis, a t-test was run and the results showed that there was a significant difference ( $t = -3,73353$   $p = ,001408$ ), since  $p < 0.05$ .

**Table 4.3**  
**Group A (Test 2vs Test 3)**  
T-test for dependent samples

Variable	Mean	Std.Dv.	N	t	df	p
Group A 2	66,50000	9,880869	20	-3,73353	19	,001408
Group A 3	78,25000	8,155560				

At 0,05 level of significance

**A.2 The following hypothesis will try to answer Research Question 2**

**Group B**

**Hypothesis 0:** There is no significant difference between the mean scores of group B's first and second tests.

**Hypothesis 1:** There is a significant difference between the mean scores of group B's first and second tests.

In order to test the null hypothesis, a t-test was run and the results showed that there was not a significant difference ( $t = -2,23802$   $p = ,037392$ ), since  $p < 0.05$ .

**Table 4.4**  
**Group B (Test 1vs Test 2)**  
T-test for dependent samples

Variable	Mean	Std.Dv.	N	t	df	p
Group B 1	69,50000	10,50063	20	-2,23802	19	,037392
Group B 2	75,00000	5,12989				

At 0,05 level of significance

**Hypothesis 0:** There is no significant difference between the mean scores of group B's first and third tests.

**Hypothesis 1:** There is a significant difference between the mean scores of group B's first and third tests.

In order to test the null hypothesis, a t-test was run and the results showed that there was a significant difference ( $t = -332745$   $p = ,003539$ ), since  $p < 0.05$ .

**Table 4.5**  
**Group B (Test 1vs Test 3)**  
T-test for dependent samples

Variable	Mean	Std.Dv.	N	t	df	p
Group B 1	69,50000	10,50063	20	-3,32745	19	,003539
Group B 3	78,50000	9,88087				

At 0,05 level of significance

**Hypothesis 0:** There is no significant difference between the mean scores of group B's second and third tests.

**Hypothesis 1:** There is a significant difference between the mean scores of group B's second and third tests.

In order to test the null hypothesis, a t-test was run and the results showed that there was not a significant difference ( $t = -1,67711$   $p = ,109897$ ), since  $p < 0.05$ .

**Table 4.6**  
**Group B (Test 2vs Test 3)**  
T-test for dependent samples

Variable	Mean	Std.Dv.	N	t	df	p
Group B 2	75,00000	5,129892	20	-1,67711	19	,109897
Group B 3	78,50000	9,88087				

At 0,05 level of significance

**A.3 The following hypothesis will try to answer Research Question 3**

**Group C**

**Hypothesis 0:** There is no significant difference between the mean scores of group C's first and second tests.

**Hypothesis 1:** There is a significant difference between the mean scores of group C's first and second tests.

In order to test the null hypothesis, a t-test was run and the results showed that there was a significant difference ( $t = -3,24893$   $p = ,004224$ ), since  $p < 0.05$ .

**Table 4.7**  
**Group C (Test 1vs Test 2)**  
T-test for dependent samples

Variable	Mean	Std.Dv.	N	t	df	p
Group C 1	64,50000	10,99043	20	-3,24893	19	,004224
Group C 2	69,50000	6,86333				

At 0,05 level of significance

**Hypothesis 0:** There is no significant difference between the mean scores of group C's first and third tests.

**Hypothesis 1:** There is a significant difference between the mean scores of group C's first and third tests.

In order to test the null hypothesis, a t-test was run and the results showed that there was a significant difference ( $t = -6,24228$   $p = ,000005$ ), since  $p < 0.05$ .

**Table 4.8**  
**Group C (Test 1vs Test 3)**  
T-test for dependent samples

Variable	Mean	Std.Dv.	N	t	df	p
Group C 1	64,50000	10,99043	20	-6,24228	19	,000005
Group C 3	81,00000	11,19210				

At 0,05 level of significance

**Hypothesis 0:** There is no significant difference between the mean scores of group C's second and third tests.

**Hypothesis 1:** There is a significant difference between the mean scores of group C's second and third tests.

In order to test the null hypothesis, a t-test was run and the results showed that there was a significant difference ( $t = -4,94520$   $p = ,000090$ ), since  $p < 0.05$ .

**Table 4.9**

**Group C (Test 2vs Test 3)**

T-test for dependent samples

Variable	Mean	Std.Dv.	N	t	df	p
Group C 2	69,50000	6,86333	20	-4,94520	19	,000090
Group C 3	81,00000	11,19210				

At 0,05 level of significance

In summary, t-test for dependent samples showed a significant difference for Group A, Group B, and Group C.

**4.1.3. Comparison of the tests between the groups:**

**The following hypothesis will try to answer Research**

**Question 4**

**1<sup>st</sup> test:**

**Hypothesis 0:** There is a significant difference between the mean scores of group A's first test and the mean scores of group B's first test.

**Hypothesis 1:** There is not a significant difference between the mean scores of group A's first test and the mean scores of group B's first test.

In order to test the null hypothesis, a t-test was run and the results showed that there was a significant difference ( $t = -5,20153$   $p = ,000007$ ), since  $p < 0,05$ .

**Table 4.10**

**First Test ( Group A vs Group B)**

T-test for independent samples

Variable	Mean Group A	Mean Group B	St. Dv. Group A	St. Dv. Group B	t	df	p
First Test	53,25000	69,50000	9,215976	10,50063	-5,20153	38	,000007

At 0,05 level of significance

**Hypothesis 0:** There is a significant difference between the mean scores of group A's first test and the mean scores of group C's first test.

**Hypothesis 1:** There is not a significant difference between the mean scores of group A's first test and the mean scores of group C's first test.

In order to test the null hypothesis, a t-test was run and the results showed that there was a significant difference ( $t = -3,50772$   $p = ,001179$ ), since  $p < 0.05$ .

**Table 4.11**  
**First Test ( Group A vs Group C)**  
T-test for independent samples

Variable	Mean Group A	Mean Group C	St. Dv. Group A	St. Dv. Group C	t	df	p
First Test	53,25000	64,50000	9,215976	10,99043	-3,50772	38	,001179

At 0,05 level of significance

**Hypothesis 0:** There is a significant difference between the mean scores of group B's first test and the mean scores of group C's first test.

**Hypothesis 1:** There is not a significant difference between the mean scores of group B's first test and the mean scores of group C's first test.

In order to test the null hypothesis, a t-test was run and the results showed that there was not a significant difference ( $t = 1,471057$   $p = ,149509$ ), since  $p > 0.05$ .

**Table 4.12**  
**First Test ( Group B vs Group C)**  
T-test for independent samples

Variable	Mean Group B	Mean Group C	St. Dv. Group B	St. Dv. Group C	t	df	p
First Test	69,50000	64,50000	10,50063	10,99043	1.471057	38	,149509

At 0,05 level of significance

**B.2 The following hypothesis will try to answer Research Question 5**

2<sup>nd</sup> test:

**Hypothesis 0:** In terms of the 2nd test, there is not a significant difference between the mean scores of group A and group B.

**Hypothesis 1:** In terms of the 2nd test, there is a significant difference between the mean scores of group A and group B.

In order to test the null hypothesis, a t-test was run and the results showed that there was a significant difference ( $t = -3,41441$   $p = ,006353$ ), since  $p < 0.05$ .

**Table 4.13**  
**Second Test ( Group A vs Group B)**  
T-test for independent samples

Variable	Mean Group A	Mean Group B	St. Dv. Group A	St. Dv. Group B	t	df	p
Second Test	70,00000	75,00000	9,880869	5,129892	-3,41441	38	,001533

At 0,05 level of significance

**Hypothesis 0:** In terms of the 2nd test, there is not a significant difference between the mean scores of Group A and C.

**Hypothesis 1:** In terms of the 2nd test, there is a significant difference between the mean scores of Group A and C.

In order to test the null hypothesis, a t-test was run and the results showed that there was not a significant difference ( $t = -1,11518$   $p = ,271775$ ), since  $p > 0.05$ .

**Table 4.14**  
**Second Test ( Group A vs Group C)**  
T-test for independent samples

Variable	Mean Group A	Mean Group B	St. Dv. Group A	St. Dv. Group B	t	df	p
Second Test	70,00000	69,50000	9,880869	6,863327	-1,11518	38	,271775

At 0,05 level of significance

**Hypothesis 0:** In terms of the 2nd test, there is not a significant difference between the mean scores of Group B and C..

**Hypothesis 1:** In terms of the 2nd test, there is a significant difference between the mean scores of Group B and C.

In order to test the null hypothesis, a t-test was run and the results showed that there was a significant difference ( $t= 2,870565$   $p=,006661$ ), since  $p<0.05$ .

**Table 4.15**  
**Second Test ( Group B vs Group C)**  
T-test for independent samples

Variable	Mean Group B	Mean Group C	St. Dv. Group B	St. Dv. Group C	t	df	p
Second Test	75,00000	70,00000	5,129892	6,863327	2,870565	38	006661

At 0,05 level of significance

**B.3 The following hypothesis will try to answer Research Question 6**

3<sup>rd</sup> test

**Hypothesis 0:** In terms of the 3<sup>rd</sup> test, there is a significant difference between the mean scores of Group A and B.

**Hypothesis 1:** In terms of the 3<sup>rd</sup> test, there is not a significant difference between the mean scores of Group A and B.

In order to test the null hypothesis, a t-test was run and the results showed that there was not a significant difference ( $t= -087265$   $p=,930919$ ), since  $p>0.05$ .

**Table 4.16**  
**Third Test ( Group A vs Group B)**  
T-test for independent samples

Variable	Mean Group A	Mean Group B	St. Dv. Group A	St. Dv. Group B	t	df	p
Third Test	78,25000	78,50000	8,155560	9,880869	-,087265	38	,930919

At 0,05 level of significance



**Hypothesis 0:** In terms of the 3rd test, there is a significant difference between the mean scores of group A and C.

**Hypothesis 1:** In terms of the 3rd test, there is not a significant difference between the mean scores of Group A and C.

In order to test the null hypothesis, a t-test was run and the results showed that there was not a significant difference ( $t = -0,888076$   $p = 0,380087$ ), since  $p > 0,05$ .

**Table 4.17**  
**Third Test ( Group A vs Group C)**  
T-test for independent samples

Variable	Mean Group A	Mean Group C	St. Dv. Group A	St. Dv. Group C	t	df	p
Third Test	78,25000	81,00000	8,155560	11,19210	-,888076	38	,380087

At 0,05 level of significance

**Hypothesis 0:** In terms of the 3rd test, there is a significant difference between the mean scores of Group B and C.

**Hypothesis 1:** In terms of the 3<sup>rd</sup> test, there is not a significant difference between the mean scores of Group B and C.

In order to test the null hypothesis, a t-test was run and the results showed that there was not a significant difference ( $t = -0,748868$   $p = 0,458548$ ), since  $p > 0,05$ .

**Table 4.18**  
**Third Test ( Group B vs Group C)**  
T-test for independent samples

Variable	Mean Group B	Mean Group C	St. Dv. Group B	St. Dv. Group C	t	df	p
Third Test	78,50000	81,00000	9,880869	11,19210	-,748868	38	,458548

At 0,05 level of significance

#### 4.1.4. T-test for Dependent Samples:

##### A. Comparison of pre and post placement scores within the groups:

The following hypothesis will try to answer Research Question 7

##### Group A

**Hypothesis 0:** There is no significant difference between the mean scores of group A's pre and post placement tests.

**Hypothesis 1:** There is a significant difference between the mean scores of group A's pre and post placement tests.

In order to test the null hypothesis, a t-test was run. The results of the t-test for dependent samples showed that there was a significant difference between the scores of Open-Ended group's pre and post placement tests, at the 0.05 level of significance ( $t = -29.7884$   $p = 0,0000$ ), since  $p < 0.05$ .

Table 4.19

Pre Placement Test vs Post Placement Test ( Group A)  
T-test for dependent samples

Variable	Mean Group A	St. Dv. Group A	N	t	Df	p
Pre-placement	39,6500	2,758241	20	-29,7844	19	,000000
Post-placement	57,35000	4,158378				

At 0,05 level of significance

##### Group B

**Hypothesis 0:** There is not a significant difference between the mean scores of Group B's pre and post placement tests.

**Hypothesis 1:** There is a significant difference between the mean scores of group B's pre and post placement tests.

In order to test the null hypothesis, a t-test was run. The results of the t-test for dependent samples showed that there was a significant difference between the scores of Group B's pre and post placement tests, at the 0.05 level of significance ( $t = -23.5509$   $p = 0,0000$ ), since  $p < 0.05$ .

**Table 4.20**  
**Pre Placement Test vs Post Placement Test ( Group B)**  
T-test for dependent samples

Variable	Mean Group B	St. Dv. Group B	N	t	Df	p
Pre-placement	39,60000	2,760625	20	-23,5509	19	,000000
Post-placement	56,60000	4,913997				

At 0,05 level of significance

### Group C

**Hypothesis 0:** There is not a significant difference between the mean scores of group C's pre and post placement tests.

**Hypothesis 1:** There is a significant difference between the mean scores of group C's pre and post placement tests.

In order to test the null hypothesis, a t-test was run. The results of the t-test for dependent samples showed that there was a significant difference between the scores of Group C's pre and post placement tests, at the 0.05 level of significance ( $t = -39.9742$   $p = 0,0000$ ), since  $p < 0.05$ .

**Table 4.21**  
**Pre Placement Test vs Post Placement Test ( Group C)**  
T-test for dependent samples

Variable	Mean Group C	St. Dv. Group C	N	t	Df	p
Pre-placement	40,15000	3,265046	20	-,399742	19	,000000
Post-placement	58,60000	4,185187				

At 0,05 level of significance

In order to have inter-rater reliability, group A's tests were scored by two different teachers. One of the scorers was the researcher, the other one was an instructor from the preparatory school of Osmangazi University. To find out whether there was a difference between the scorers, a t-test was run:

**Hypothesis 0:** There is not a significant difference between the scorers in the assessment of the first quiz of Group A.

**Hypothesis 1:** There is a significant difference between the scorers in the assessment of the first test of Group A.

In order to test null hypothesis a t-test was run and the results showed that there was not a significant difference between the scorers in the assessment of the first test of group A at the level of significance, ( $t=-,423614$   $p=,674236$ ) since  $p>0.05$ .

**Table 4.22**  
**Interrelationship of the scorers ( Open-ended Test 1)**  
 T-test for dependent samples

Variable	Mean	St. Dv.	t	df	p
Scorer 1	53,25000	9,215976	-,423614	38	,674236
Scorer 2	54,50000	9,445132			

At 0,05 level of significance

**Hypothesis 0:** There is not a significant difference between the scorers in the assessment of the second test of Group A.

**Hypothesis 1:** There is a significant difference between the scorers in the assessment of the second test of Group A.

In order to test null hypothesis a t-test was run and the results showed that there was not a significant difference between the scorers in the assessment of the second test of group A at the level of significance, ( $t=-,164166$   $p=,870471$ ) since  $p>0.05$ .

**Table 4.23**  
**Interrelationship of the scorers ( Open-ended Test 2)**  
 T-test for dependent samples

Variable	Mean	St. Dv.	t	df	p
Scorer 1	66,50000	9,880869	-,164166	38	,870471
Scorer 2	67,00000	9,375219			

At 0,05 level of significance

**Hypothesis 0:** There is not a significant difference between the scorers in the assessment of the third test of group A.

**Hypothesis 1:** There is a significant difference between the scorers in the assessment of the third test of group A.

In order to test null hypothesis a t-test was run and the results showed that there was not a significant difference between the scorers in the assessment of the third test of group A at the level of significance, ( $t=-,276852$   $p=,783394$ ) since  $p>0.05$ .

**Table 4.24**  
**Interrelationship of the scorers ( Open-ended Test 3)**  
T-test for dependent samples

Variable	Mean	St. Dv.	t	df	p
Scorer 1	78,25000	8,155560	,276852	38	,783394
Scorer 2	77,50000	8,958971			

At 0,05 level of significance

In summary, a t-test for dependent samples did not show a significant difference.

## 5. DISCUSSIONS AND CONCLUSIONS

### 5.1. Summary

In this study, the effect of test method, on reading comprehension test scores was examined. The aim of this study was to find out whether different testing methods yield different scores in the assessment of reading comprehension,( and whether there was any difference among three reading comprehension test methods: Open-Ended, Multiple-Choice, True-False).

In order to do this, three intermediate level groups were formed via Michigan Placement Test. After the groups were formed, each group was given the same reading comprehension passage with a different testing method. First group ( group A) took three tests with open-ended method. The second group (group B) took three tests with multiple-choice method. The third group (group C) took three tests with true-false method. The scores obtained from the tests were compared and correlated within and between the groups. At the end of the study, the students were given the Michigan Placement Test as post-placement test to see the improvement in their linguistic level.

The results of the study were examined in three dimensions: the analysis of the test scores within groups and the analysis of test scores between the groups, the analysis of pre and post placement test scores within groups.

Table 5.1

The Overall Mean Scores of the Tests

	Group A (OE Method) Mean	Group B (MC Method) Mean	Group C (TF Method) Mean
First Test	53,25	69,5	64,5
Second Test	66,5	75	69,5
Third Test	78,25	78,5	81

In terms of first dimension: The analysis of the test scores for group A( m for the 1<sup>st</sup> test: 53. 2; m for the 2<sup>nd</sup> test: 66.5; m for the 3<sup>rd</sup> test: 78.2) showed that students had a progress in each test and a significant difference was observed between the scores.

The analysis of the test scores for group B( m for the 1<sup>st</sup> test: 69.5; m for the 2<sup>nd</sup> test: 75; m for the 3<sup>rd</sup> test: 78.5) showed that students had a progress in each test and a significant difference was observed between the scores of the first test and the second one, the first test and the third one. The results of the second and third tests were slightly different.

The analysis of the test scores for group C (m for the 1<sup>st</sup> test: 64.5; m for the 2<sup>nd</sup> test: 69.5; m for the 3<sup>rd</sup> test: 81) showed that students had a progress in each test and a significant difference was observed between the scores of the first test and the second one, the first test and the third one, the second test and the third one.

In terms of the second dimension: The results of the first test showed that, the students of group B, who took the reading comprehension test with Multiple-Choice method, got higher scores ( m= 69) than the other groups. The second highest score was taken by the students of group C ( m= 64.5), who took the tests with True-False method. The

m =53 was taken by the students of group A, who took the tests with OE method. As a result a significant difference was observed between the mean scores of group A and B, group A and C. However, there was not a significant difference between group B and C. When the results of the second test were taken, it was seen that students of group B took the highest score (m=75). Group C took the second highest score ( m=69). The m=66.5 was taken by the students of group A. As a result there was a significant difference between group A and B, group B and C but there was not a significant difference between group A and C.

And finally, the third test results revealed that the students of group C got higher scores (m = 81) than the students of group A ( m = 78.5) and group B ( m = 78.5). As a result no significant difference was observed between the groups.

In terms of the third dimension: The results of pre and post placement test scores for each group(pre-placement test mean scores for group A = 39.6 , for group B = 39.6 , for group C = 40.1 ; post-placement test mean scores for group A = 57.3 , for group B = 56.6 , for group C = 58.6 ) suggested that significant difference was observed between pre and post placement test scores.

## 5.2 Discussion

Do different testing methods yield different scores in the assessment of reading comprehension? The results of this study show that this question is a matter of proficiency. In this study, subjects performed better on the Multiple-choice tests than on the Open-Ended and True-false tests only in the first test, which was given when the proficiency level of students were low. In terms of the 2<sup>nd</sup> tests' scores it was seen that the difference between the scores decreased and finally, students got similar results in



the third tests, which was given at the end of the study when students had improvement in terms of English proficiency. Thus it can be said that testing method became an insignificant factor by the improvement of proficiency. In other words, as the proficiency improved different test methods measured the trait in the same manner. With these results research question 8 can be answered and it can be said that there is an effect of proficiency on test type effect.

These findings closely resemble those of Stathmann(1979), whose subjects also got the same proportion of Multiple-choice and Short-answer questions correct in a test which included both Multiple-choice and Short-answer questions equally.

Furthermore, these findings are supported by a study of Samson(1983, in Shohamy 1984), which he found out that Multiple-choice, Open-ended and, Summary writing testing methods measured the same trait and there were no major differences.

Another study carried out by Bormuth (1967, in Bensoussan 1984) also supports the findings of this study. In his study, the subjects got equivalent close and multiple choice comprehension test percentage scores from the same text.

In the present study, students got higher scores on the multiple choice tests at the beginning of the study. Possible explanation for subjects' higher performance on the multiple choice test is "guessing". Subjects could have received high scores on the Multiple-Choice questions by guessing some of the items correctly. The Multiple-Choice items provided retrieval cues-the correct response was one of the choices- that facilitated the comprehension of passage information.

The Open-Ended test type, which students got the lowest score at the beginning of the study, provided no such retrieval cues and therefore couldn't aid the subjects who may not have been able to recall the information necessary to write an answer.

Students got higher points on True-False tests, too. In terms of the first test an insignificant difference was observed between Multiple-Choice and True-False tests. Since the students had fifty percent chance for the correct answer, guessing could be possible explanation for this result.

Subjects answering Multiple-Choice questions could have guessed some responses correctly by recognising words or phrases from the passage without having understood what they read. Subjects answering Open-Ended questions, on the other hand, did not receive such cues.

Lower mean scores of Open-Ended type tests, for the first and second tests, such as that due to lower proficiency, students could have problems as they did not have any cues as the other students got in Multiple-Choice and True-False tests. Another explanation can be their limited vocabulary and their difficulty in producing the target language. The subjects had difficulty to express what they understood as evidenced by their low scores.

Shohamy (1984) states that, doing Multiple-Choice items involves comprehension and selection while Open-Ended requires comprehension and production. Shohamy(1984) points out that low level test takers are most effected by the testing method variables, while high level test takers are hardly affected.

The results of the present study show that as subjects become more proficient in the target language, they are less likely to be affected by assessment task type. Shohamy (1984) posits that more advanced students tend to do equally well regardless of the method, because they are able to manipulate the target language.

The present research provided information on the effect of testing methods on students' scores, and it showed that testing method did not yield different scores when

the student became more proficient. An insignificant difference may be observed for low level students. When the subjects become proficient, the test method become an insignificant factor.

Looking at the overall scores of the all groups, it can be said that there is an effect of proficiency on test type effect. The more proficient they become , the less insignificant scores they get from the tests they are given with different testing methods.

The results from this study suggest that testers should be sensitive to the proficiency level of the test takers, since testing method make a difference for low level students than for more proficient ones. Shohamy (1984) suggests using multiple approaches for testing reading comprehension and not to rely on one method.

It should be noted that the aim of this study was not to find out the best method. Any single method can not be said to be the best method to assess the reading comprehension.

Since there is not one true comprehension of a text, there is no perfect test of comprehension. Testing can create additional interactions that can impede readers' ability to demonstrate what they understood. No one test can measure all aspects of reading comprehension ( Wolf, 1993).

### **5.3. Suggestions for further studies**

In this study, intermediate level EFL students were used. Further studies can be conducted with elementary, upper intermediate and advanced level students to see if there will be differences in the results depending on the level.

In further studies other testing methods, which were not included in this study, can be used. Also due to time constraints, each group took only one version of the

passage; Group A took the tests only with Open-Ended method, Group B took the tests with multiple-choice method, and Group C took the tests only with True-False method.

In further studies, each group can take the tests with three different testing methods, for e.g.: those students who take the Open-Ended version will take the Multiple-Choice version and True-False version of the same passage and vice versa.

This study was carried out in a semester time. A longitudinal research can be conducted to see if time is a significant factor in assessing reading comprehension using different testing methods.

## APPENDIX A1

### ORDINARY ASPIRIN IS TRULY A WONDER DRUG

[Aspirin is the most popular drug in the world today. Many people take aspirin when they have a headache. It is effective in relieving other pains too. This article discusses the history and uses of aspirin.]

1 Americans this year will swallow 15,000 tons of aspirin, one of the safest and most effective drugs invented by man. The most popular medicine in the world today, it is an effective pain reliever. Its bad effects are relatively mild, and it is cheap.

2 For millions of people suffering from arthritis, it is the only thing that works. Aspirin, in short, is truly the 20<sup>th</sup>-century wonder drug. It is also the second largest suicide drug and is the leading cause of poisoning among children. It has side effects that, although relatively mild, are largely unrecognised among users.

3 Although aspirin was first sold by a German company in 1899, it has been around much longer than that. Hippocrates, in ancient Greece, understood the medical value of the leaves and tree bark which today are known to contain salicylates, the chemical in aspirin. During the 19th century, there was a great deal of experimentation in Europe with this chemical, and it led to the introduction of aspirin. By 1915, aspirin tablets were available in the United States.

4. A small quantity of aspirin (two five-grain tablets) relieves pain and inflammation. It also reduces fever by interfering with some of the body's reactions. Specifically, aspirin seems to slow down the formation of the acids involved in pain and the complex chemical reactions that cause fever. The chemistry of these acids is not fully understood, but the slowing effect of aspirin is well known.

5 Aspirin is very irritating to the stomach lining, and many aspirin takers complain about upset stomach. There is a right way and a wrong way to take aspirin. The best way is to chew the tablets before swallowing them with water, but few people can stand; the bitter taste. Some people suggest crushing the tablets in milk or orange juice and drinking that.

## APPENDIX A2

A. Read the passage carefully and answer the following questions:

1. What does the article mainly discuss?
2. What does paragraph 1 describe?
3. Is aspirin always safe? Why? Why not?
4. What does paragraph 3 describe?
5. What are salicylates?
6. What do we learn from paragraph 4?
7. What is the importance of aspirin for fever ?
8. What does paragraph 5 describe?
9. Does the author of the article seem to be in favor of aspirin? Does the author of the article seem to be against the use of aspirin?
10. Can you write a title for the article which emphasizes the advantages of aspirin?

### ANSWER KEY

1. The article discusses both the good and bad things about aspirin.
2. Paragraph 1 describes the good things about aspirin.
3. No, it is not always safe. It is the second largest suicide drug and also, it is the leading cause of poisoning among children. Also, it may cause stomach ache.
4. Paragraph 3 describes the history of aspirin.
5. Salicylates are the chemical in aspirin.
6. We learn about how aspirin works in the body.
7. It slows down the reactions that cause fever.
8. Paragraph 5 describes how to take aspirin.
9. The author seems to be in favor of aspirin.
10. Any title supporting the advantages of aspirin will be acceptable.

## APPENDIX A3

A. Read the passage carefully and circle the letter next to the best answer:

1. This article discusses:
  - a. only the good things about aspirin.
  - b. both the good and bad things about aspirin.
  - c. only the bad things about aspirin.
  
2. Paragraph 1 describes:
  - a. what aspirin is made of.
  - b. the good things about aspirin.
  - c. the bad things about aspirin.
  
3. The information in paragraph 2 shows that:
  - a. aspirin can be dangerous.
  - b. aspirin is always safe.
  - c. aspirin has been around a long time
  
4. Paragraph 3 describes the ..... of aspirin.
  - a. value
  - b. uses
  - c. history
  
5. In paragraph 3, what are **salicylates**?
  - a. Leaves and tree bark.
  - b. The chemical in aspirin.
  - c. Aspirin tablets
  
6. Paragraph 4 describes:
  - a. how aspirin works in the body.
  - b. the side effects of aspirin.
  - c. how to take aspirin.
  
7. Aspirin seems to .....A..... the reactions that .....B..... fever.  
(Choose one word from each group.)  
A: speed up - slow down - stop  
B: cause - stop - understand
  
8. Paragraph 5 describes:
  - a. experimentation with aspirin.
  - b. how aspirin works in the body.
  - c. how to take aspirin.

9. The author of this article seems to be:
- a. in favor of aspirin.
  - b. against the use of aspirin.
  - c. not interested in aspirin.
10. A good title for this article would be:
- a. Aspirin: The Most Popular Medicine in the World
  - b. Why I Take Aspirin
  - c. The Side Effects of Aspirin

### ANSWER KEY

- 1.B
- 2.B
- 3.A
- 4.C
- 5.B
- 6.A
- 7.Slow down / cause
- 8.C
- 9.A
- 10.A



## APPENDIX A4

A. Read the passage carefully and decide whether they are True or False.(Write T or F in the blanks)

- \_\_\_ 1. This article discusses only the bad things about aspirin.
- \_\_\_ 2. Paragraph 1 describes what aspirin is made of.
- \_\_\_ 3. The information in paragraph 2 shows that aspirin can be dangerous.
- \_\_\_ 4. Paragraph 3 describes the value of aspirin.
- \_\_\_ 5. Salicylates are the chemical in aspirin.
- \_\_\_ 6. Paragraph 4 describes the side effects of aspirin.
- \_\_\_ 7. Aspirin seems to slow down the reactions that cause fever.
- \_\_\_ 8. Paragraph 5 describes how aspirin works in the body.
- \_\_\_ 9. The author of this article seems to be in favour of aspirin.
- \_\_\_ 10. A good title for this article would be “ The side effects of aspirin”.

### ANSWER KEY

- |      |      |      |      |       |
|------|------|------|------|-------|
| 1. F | 3. T | 5. T | 7. T | 9. T  |
| 2. F | 4. F | 6. F | 8. F | 10. F |

## APPENDIX B1

### WHAT YOU DON'T KNOW ABOUT EXERCISE

[Exercise is good for you, doctors say but most people really know very little about how to exercise properly. What do you know about exercise? Take this true-false test. You will probably be surprised

A. Exercise, everyone advises! But immediately when you try, you run into trouble. There is so much contradictory, sometimes incorrect advice that you become confused. Test yourself on the following true-false quiz. It will tell you what you need to know. It will tell you what you need to know.

1. *The best way to reduce mid-section is to do abdominal exercise.*

False. Many people believe that when specific muscles are exercised, the fatty tissues in the immediate area are "burned up". The truth is that exercise burns fat from all over the body and not from one specific area, regardless of the type of exercise. Of course if you reduce the fat through out your body, you will certainly see results around your waist line too!

2. *To maintain an adequate level of physical fitness, you need to exercise only twice a week.*

False. Studies conducted by NASA, the National Aeronautics and Space Administration, show that unexercised muscles lose strength very quickly. After 48 to 72 hours, you must use the muscles again to re-establish the good physical effects. And what does that mean to you? NASA scientists concluded that while daily exercise is most beneficial, three alternating days each week will maintain an adequate level of physical fitness.

3. *To lose weight you should always "work up a good sweat" when exercising.*

False. Sweating only lowers body temperature to prevent overheating; it does not help you reduce. You may weigh less immediately after a workout, but this is due to water loss. Once you replace the liquid, you replace the weight.

4. *You burn more calories jogging one mile than walking the same distance.*

False. You use the same amount of energy whether you walk or jog the mile, since in both cases you are moving the same weight the same distance. The speed doesn't matter. Of course, if you jog rather than walk for 30 minutes, you'll cover more distance, and therefore burn more calories.

5. *If your breathing does not return to normal within minutes after you finish exercising, you have exercised too much.*

True. Five minutes or so after exercising, your breathing should be normal, your heart shouldn't be pounding, and you shouldn't be exhausted. Beneficial exercise is not overly difficult, unpleasant, and exhausting; it is moderate, enjoyable, and refreshing.

6. *Walking is one of the best exercises.*

True. Walking helps circulation of blood throughout the body, and thus has a direct effect on your overall feeling of health.

7. *Vigorous stretching exercises keep muscles flexible.*

False. Stretching exercises (for example: twisting or bending at the waist, touching your toes) should be done slowly, allowing the muscles to relax and “let go”. Vigorous stretching makes the muscles become tighter.

8. *The minimum amount of time you should spend exercising in a day is 20 minutes.*

True. There are more than 400 muscles that attach to your skeleton. A good exercise routine should contract and stretch all these muscles, and this simply cannot be done with four or five exercises in five or ten minutes. From experience, I've found that about 20 minutes is the minimum amount of time needed for an adequate workout.

B. How long it takes you to become physically fit depends on how unfit you are when you start. If you are out of condition, you certainly can't shape up in 21 days. However, shaping up doesn't do any good unless you plan to stay in shape, and that means exercising from now on. It takes as much exercise to stay in shape as it does to get there. But the work won't seem as hard after a while because your body will be in good condition-and all the moves will seem easier.

## APPENDIX B 2

A. Read the passage carefully and answer the questions

1. What does the article mainly explain?
2. In paragraph A, “it will tell you .....” what does **it** refer to?
3. What does paragraph 1 explain?
4. To maintain an adequate level of physical fitness how often do you need to exercise according to the second paragraph?
5. According to paragraph 4, how will you use the most energy?
6. What does paragraph 5 compare?
7. What does paragraph 6 explain?
8. In paragraph, 7 why is (twisting or bending at the waist, touching your toes) in parentheses?
9. According to the last paragraph, once you become physically fit, what should you do?
10. Is the writer in favour of taking exercise? Is the writer against exercise?

### ANSWER KEY

1. The article explains important information to know about exercise.
2. It refers to the true-false quiz.
3. Paragraph 1 explains why you can't reduce just one area of your body.
4. Daily exercise is most beneficial, also it is adequate to exercise on three alternating days.
5. By covering more distance.
6. Paragraph 5 compares the results of beneficial exercise and too much exercise.
7. Paragraph 6 explains why walking is one of the best exercises.
8. Because these are examples of stretching exercises.
9. You should keep on exercising to stay fit.
10. The writer is in favour of taking exercise.

### APPENDIX B 3

A. Read the passage carefully and answer the following questions

1. This article explains:

- a. the difficulties of exercising.
- b. why exercise is good for your health.
- c. important information to know about exercise.

2. In paragraph A, "it will tell you. . .," it refers to:

- a. the true-false quiz.
- b. exercise.
- c. the advice.

3. Paragraph 1 explains:

- a. how to reduce one specific area of your body.
- b. why you can't reduce just one area of your body.
- c. how to do abdominal exercises.

4. It is \_\_\_\_\_ A \_\_\_\_\_ to exercise on three \_\_\_\_\_ B \_\_\_\_\_ days each week.

(Choose one word from each group.)

A: not enough - adequate - too much

B: alternating - daily - beneficial

5. According to paragraph 4, you will use the most energy if you:

- a. jog for 30 minutes.
- b. walk for 30 minutes.
- c. run quickly for 30 minutes.

6. Paragraph 5:

- a. says you should exercise for five minutes a day.
- b. compares the results of beneficial exercise and too much exercise.
- c. gives examples of beneficial exercise.

7. Paragraph 6 explains:

- a. why you feel healthy.
- b. how blood circulates throughout the body.
- c. why walking is one of the best exercises.

8. In paragraph 7, why is "(twisting or bending at the waist, touching your toes)" in parentheses?

- a. Because these exercises should be done slowly.
- b. Because these are examples of stretching exercises.
- d. Because these are not important exercises.

9. According to the last paragraph, once you become physically fit:
- you will stay that way naturally.
  - you must keep on exercising to stay fit
  - exercise will be hard work.
10. The author of this article probably:
- exercises regularly.
  - doesn't like to exercise
  - exercises 10 minutes a day.

#### ANSWER KEY

- C
- A
- B
- Adequate/ alternating
- A
- B
- C
- B
- B
- A

## APPENDIX B.4

A. Read the passage carefully and decide whether they are true or false.( Write T or F )

- \_\_\_ 1. This article explains the difficulties of exercising.
- \_\_\_ 2. In paragraph A, “it will tell you” it refers to the advice.
- \_\_\_ 3. Paragraph 1 explains why you can not reduce just one area of your body.
- \_\_\_ 4. It is adequate to exercise on three alternating days each week.
- \_\_\_ 5. According to paragraph 4, you will use the most energy if you walk for 30 minutes.
- \_\_\_ 6. Paragraph 5 compares the results of beneficial exercise and too much exercise.
- \_\_\_ 7. Paragraph 6 explains how blood circulates throughout the body.
- \_\_\_ 8. In paragraph 7, “twisting or bending at the waist, touching your toes” is in parentheses because these are examples of stretching exercises.
- \_\_\_ 9. According to the last paragraph, once you become physically fit you will stay that way naturally.
- \_\_\_ 10. The author of this article probably does not like to exercise.

### ANSWER KEY

- |     |     |     |     |      |
|-----|-----|-----|-----|------|
| 1.F | 3.T | 5.F | 7.F | 9.F  |
| 2.F | 4.T | 6.T | 8.T | 10.F |

## APPENDIX C 1

### HOW TO GIVE A GOOD SPEECH

[We are all called upon to make a speech at some point in life, but most of us don't do a very good job. This article gives some suggestions on how to give an effective speech.]

1 So, you have to give a speech-and you're terrified. You get nervous, you forget what you want to, say, you stumble over words, you talk too long, and you bore your audience. Later you think, "Thank goodness its over. I'm just no good at public speaking. I hope I never have to do that again."

2 Cheer up! It doesn't have to be that bad. Here are some simple steps to take the pain out of speech-making. First of all, it is important to plan. Do your homework. Find out everything you can about your subject. And, at the same time, find out as much as you can about your audience. Who are they? What do they know about your subject? Do they have a common interest? Why are they coming to hear you speak? Put yourself in their shoes as you prepare your speech.

3 Ask yourself the purpose of your speech. What is the occasion? Why are you speaking? Are you introducing another speaker? Moderating a discussion? Giving a lecture? Convincing someone? There are many possible speaking roles, and each one has its own special characteristics. Make sure you know into which category you fit. Don't spoil your speech by confusing one speaking role with another.

4 Let us suppose that you have been asked to introduce the main speaker at a conference. First, find out the most important and interesting things about the speaker. Then, summarise this information in a few remarks. It is all right to tell a joke or an anecdote if it is in good taste and will not embarrass the speaker. And, most important, be brief. Remember, you are not the main speaker; you are *introducing* the main speaker.

5 If you are a moderator, you should begin by giving a quick introduction of the people on the panel. After that, you should try to keep the discussion running smoothly, and you should try to focus on the connections between speakers. Keep yourself in the background. Don't talk too much, and don't interrupt the panelists. Be tactful and be considerate.

6 If you are giving a lecture or explaining an idea, gather as many facts as you can on your subject. Spend plenty of time doing your research. Then spend plenty of time organising your material so that your speech is clear and easy to follow. Use as many examples as possible, and use pictures, charts, and graphs if they will help you make your points more clearly. Never forget your audience. Don't talk over their heads, and don't talk down to them. Treat your audience with respect. They will appreciate your thoughtfulness.

7 If you are trying to sell something, you will need to convince your audience. Do you want them to vote for Candidate A? Are you offering them a new, improved toothbrush? This kind of speech is usually dramatic, but here too, you must do your research and know your facts.



- 8 When you are making your speech, try to relax. Speak slowly and clearly and look at people in your audience. Use simple vocabulary and expressions whenever possible. Pause for a few seconds now and then to give your audience a chance to think about what you have said. Make sure that everyone in the room can hear you. If it is a large room or an auditorium, you will probably have to use a micro-phone.
- 9 Just remember: be prepared. Know your subject your audience, and the occasion. Be brief. Say what you have to say and then stop. And be" yourself. Let your personality come through so that you make person to person contact with your audience.
- 10 If you follow these simple steps, you will see that you don't have to be afraid of public speaking in fact you may find the experience so enjoyable that you volunteer to make more speeches! You are not convinced yet? Give it a try and see what happens.

## APPENDIX C.2

- A. Read the passage carefully and answer the following questions.
1. What is the main idea of the article?
  2. According to paragraph 1, how do many people feel about giving a speech?
  3. What does paragraph 2 describe?
  4. “ put yourself in their shoes (In paragraph 2), what does the writer want you to do?
  5. What does paragraph 3 explain?
  6. In paragraph 4, why are the words “ you” and “introducing” in italics?
  7. If you are a moderator, what should you do?
  8. What should you do to convince your audience?
  9. What does paragraph 8 explain?
  10. If the author of this article were speaking to you instead of writing, what type of speaker would he/she be ?

## ANSWER KEY

1. The main idea of the article is you can improve your speaking ability.
2. Many people are afraid of giving a speech.
3. Paragraph 2 gives some suggestions for how to make a better speech.
4. The writer wants you to try to imagine how they think or feel.
5. Paragraph 3 explains that there are different kinds of speeches.
6. They are in italics for emphasis.
7. You should act as a bridge between the speakers, you should keep yourself in the background.
8. You should do your research, know your facts, and your speech should be dramatic.
9. Paragraph 8 summarises some basics for how to make a good speech.
10. Lecturer .

### APPENDIX C.3

A. Read the passage carefully and circle the letter next to the best answer.

1. The main idea of this article is:
  - a. you can improve your speaking ability.
  - b. a poor speaker can never change.
  - c. always make a short speech.
  
2. Paragraph 1 implies, but does not directly say, that:
  - a. many people talk too long
  - a. many people are afraid of giving a speech.
  - b. many people are happy to give a speech.
  - c. many people are afraid of giving a speech.
  
3. Paragraph 2:
  - a. says it is difficult to
  - b. gives some suggestions for how to make a better speech.
  - c. does not give any suggestions for improvement.
  
4. In paragraph 3, what is the correct beginning for the phrases "Moderating a discussion? Giving a lecture? Convincing someone?"
  - a. Are you. . .
  - b. The occasion is. . .
  - c. What is. . .
  
5. Paragraph 3 explains that:
  - a. all speeches are similar.
  - b. there are different kinds of speeches.
  - c. a successful speaker is always dramatic.
  
6. In paragraph 4, why are the words "you" and "introducing" in italics?
  - a. Because they are not important.
  - b. To make them easier to read.
  - c. For emphasis.
  
7. Paragraph 5 suggests that if you are a moderator:
  - a. you should talk about yourself.
  - b. you should act as a "bridge" between speakers.
  - c. you should give a very long speech.
  
8. According to paragraph 7, which of the following would be a speech to convince you of something?
  - a. How to Bake Bread
  - b. An Evening to Honour Senator Smith's Long Career
  - c. Why You Should Live in Florida

9. Paragraph 8:
- a. explains how hard it is to make a speech.
  - b. gives examples of a lecturer's role.
  - c. summarises some basics for how to make a good speech.
10. If the author of this article were speaking to you instead of writing, what type of speaker would he / she be?
- a. Moderator
  - b. Lecturer
  - c. Introducer

#### ANSWER KEY

- 1.A
- 2.C
- 3.B
- 4.A
- 5.B
- 6.C
- 7.B
- 8.C
- 9.C
- 10.B

#### APPENDIX C.4

A. Read the passage carefully and decide whether the following statements are True or False. ( Write T or F in the blanks)

- \_\_\_ 1. According to the article “ a poor speaker can never change”.
- \_\_\_ 2. Paragraph 1 implies but does not directly say that many people are afraid of giving a speech.
- \_\_\_ 3. Paragraph 2 does not give any suggestions for improvement.
- \_\_\_ 4. “ Put yourself in their shoes” ( in paragraph 2). With this statement the writer wants you to try on their shoes to see if they fit you.
- \_\_\_ 5. Paragraph 3 explains that all speeches are similar.
- \_\_\_ 6. In paragraph 4, the words “ you” and “ introducing” are in italics for emphasis.
- \_\_\_ 7. If you are a moderator you should act as a bridge between speakers.
- \_\_\_ 8. You should convince your reader if you are giving a lecture.
- \_\_\_ 9. Paragraph 8, summarises some basics for how to make a good speech.
- \_\_\_ 10. If the author of this article were speaking to you instead of writing, he/she would be an introducer.

#### ANSWER KEY

- |     |     |     |     |      |
|-----|-----|-----|-----|------|
| 1.F | 3.F | 5.F | 7.T | 9.T  |
| 2.T | 4.F | 6.T | 8.F | 10.F |

APPENDIX D

GROUPS' FIRST, SECOND, THIRD TEST and  
PRE&POST PLACEMENT TEST SCORES

	Group A pre- placement	Group A Test 1	Group A Test 2	Group A Test 3	Group A post placement	Group B pre- placement	Group B Test 1	Group B Test 2	Group B Test 3	Group B post placement	Group C pre- placement	Group C Test 1	Group C Test 2	Group C Test 3	Group C post placement
1	37	50	65	80	55	41	70	80	90	64	39	70	70	80	58
2	38	45	65	85	59	38	70	80	70	58	44	80	80	80	64
3	38	40	50	70	61	36	60	70	80	56	45	60	60	70	67
4	39	60	70	85	60	43	60	70	80	60	37	60	70	100	55
5	43	65	70	90	62	39	40	80	70	55	35	40	60	90	59
6	43	65	65	80	60	40	60	70	90	56	40	50	60	60	55
7	44	60	55	75	62	45	70	80	90	62	42	70	70	80	60
8	37	50	70	90	55	42	80	70	80	61	38	60	70	70	55
9	36	45	40	85	57	38	80	80	100	57	43	60	70	90	58
10	38	55	55	85	55	37	70	70	70	55	36	60	60	80	55
11	41	60	70	80	60	41	70	80	70	58	39	60	70	80	58
12	43	70	75	80	61	43	70	70	80	62	40	70	70	90	60
13	45	65	75	75	65	37	70	70	70	61	44	80	80	100	64
14	38	60	65	85	58	40	70	80	80	60	38	60	70	70	55
15	37	45	75	70	59	35	80	80	80	55	41	70	70	80	62
16	39	40	65	65	55	36	60	70	70	58	43	60	70	80	61
17	42	50	70	80	57	43	70	80	80	63	37	80	60	80	57
18	40	45	75	60	58	41	80	70	70	62	42	80	80	90	60
19	37	45	75	70	59	38	70	70	60	59	45	50	80	90	64
20	38	50	80	75	55	39	90	80	90	60	35	60	70	60	55

## BIBLIOGRAPHY

- Aksu, M. EDS 330. **Measurement and Evaluation (Lecture Notes)**. METU. 1997.
- Alderson, J.C. "The Close Procedure and Proficiency in English as a Foreign Language". **Tesol Quarterly** 13: 219-227. 1979.
- Baker, D. **Language Testing: A Critical Survey and Practical Guide**. London. 1989.
- Basanta, C.P. "Coming to Grips with Progress: Testing Some Guidelines for its design". **Teaching English Forum** Vol.33:55-56. 1995.
- Bensoussan, M. "A Comparison of Cloze and Multiple Choice Reading Comprehension Tests of English as a Foreign Language". **Language Testing**, 1(1):101-104. 1984.
- Cohen, A.D. **Testing Language Ability in the Classroom**. Rowley, Massachusetts: Newbury House Publishers. 1980.
- Davey, B. "Assessing Comprehension: Selected Instructions of Task and Reader". **The Reading Teacher** 42: 694-697. 1989.
- Enginarlar, H. "A Quantitative and Qualitative Comparison of Three Techniques of Grading ESL/ EFL Essays". **Journal of Human Sciences** 10 (1): 23-45. 1991.
- Gordon, C.M, David Hanauer "The Interaction Between Task and Meaning Construction in EFL Reading Comprehension Tests". **Tesol Quarterly** Vol.29:2. 1995.
- Heaton, J.B. **Writing English Language Tests**. Hong Kong: Longman Group Ltd. 1988.
- Hughes, A. **Testing for Language Teachers**. Cambridge. Cambridge University Press. 1989.
- Kendall, J. R, et all. "Which Comprehension? Artifacts in the Measurement of Reading Comprehension". **Journal of Educational Research** 73: 233-236. 1980.
- Killoran, I. "In Defense of the Multiple Choice Question". **Social Education** 56(2):106-108. 1992.
- Madsen, H.S. **Techniques in Testing**. Oxford University Press. 1983.

Markstein, L and Hirasawa .L. **Developing Reading Skills. Int.**  
USA: Newbury House Publishers. 1981.

Mason, V.W. "Using Multiple Choice Tests to Promote Homogeneity of Class Ability Levels in Large EGP and ESP Programs".  
**System** 12(3): 263-271 1984.

Nevo, N. "Test-taking Strategies on a Multiple-Choice Test of Reading Comprehension." **Language Testing** 6(2):199-215. 1989.

Porter, D. "Cloze Procedure and Equivalence". **Language Learning** 28: 333-341.1978

Rivers, W.M. and Temperley, M.S. **A Practical Guide to the teaching of English as a Second or Foreign Language.** Oxford: Oxford University Press. 1978.

Shohamy, E. "Does the Testing Method Make a Difference? The Case of Reading Comprehension". **Language Testing**, 1:147-170. 1984.

Stathmann, S. "Testing Reading Comprehension: Multiple-choice vs Short-answer questions". **ELT Journal** 33(4): 304-307. 1979

Weir, C.J. **Communicative Language Testing.** London: Prentice Hall International Ltd.1990

Wolf, D.F. "Issues in Reading Comprehension Assessment: Implications for the development of Research Instruments and Classroom Tests". **Foreign Language Annals** Vol.26:3. 1993.

Wolf, D.F. "A comparison of Assessment Tasks Used to Measure Foreign Language Reading Comprehension". **The Modern Language Journal.** 77:473-489. 1993.