

**REPUBLIC OF TURKEY  
ÇANAKKALE ONSEKİZ MART UNIVERSITY  
INSTITUTE OF SOCIAL SCIENCES  
DEPARTMENT OF ENGLISH LANGUAGE TEACHING**

**STUDENTS' AND TEACHERS' PERCEPTIONS OF THE TEST  
TECHNIQUES USED TO ASSESS LANGUAGE SKILLS  
AT UNIVERSITY LEVEL**

**MA THESIS**

**Supervisor  
Prof. Dr. Dinçay KÖKSAL**

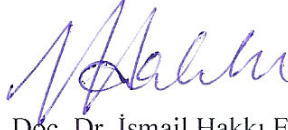
**Submitted by  
Kürşat CESUR**

**Çanakkale-2008**

Sosyal Bilimler Enstitüsü Müdürlüğüne  
Kürşat CESUR'a ait  
'Students' and Teachers' Perceptions of the Test Techniques  
Used to Assess Language Skills at University Level'  
Adlı çalışma, jürimiz tarafından  
Yabancı Diller Eğitimi Anabilim Dalı İngilizce Öğretmenliği Programında  
YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.



Başkan Prof. Dr. Dinçay KÖKSAL  
Akademik Ünvanı, Adı Soyadı (Danışman)



Üye Yrd. Doç. Dr. İsmail Hakkı ERTEN  
Akademik Ünvanı, Adı Soyadı

Üye Yrd. Doç. Dr. Aysun YAVUZ  
Akademik Ünvanı, Adı Soyadı



Üye Yrd. Doç. Dr. Cevdet YILMAZ  
Akademik Ünvanı, Adı Soyadı



Üye Yrd. Doç. Dr. İlke EVİN GENÇEL  
Akademik Ünvanı, Adı Soyadı

## ABSTRACT

This study aims (a) to find out the students' and the instructors' perceptions of the Compulsory English Language Course exams used to assess language performance at Çanakkale Onsekiz Mart University (COMU). It further aims (b) to determine what other objective test techniques can also be used in these exams in addition to the multiple-choice test technique by taking all the students' and the instructors' opinions into consideration.

Quantitative research methodology was used in this descriptive study. In the light of the literature; in order to achieve the aims stated above, two questionnaires were designed by the researcher and administered to 367 students and 33 instructors. After analyzing the internal consistency of the items in the questionnaires, the researcher found acceptable Alpha reliability values both for the students' questionnaire and for the instructors' questionnaire. Data from the students and instructors were collected by using these questionnaires. Instructors' questionnaire was administered to the instructors who had worked or were still working as the instructors of 'Compulsory English Language Course' at COMU. The students who involved in the study were all in their second years at the university and they all had the "Compulsory English Language Course" the year before the study was conducted.

The data obtained through the questionnaires were analyzed via Descriptive Statistics, One-way ANOVA, Independent Samples T-Test, Cronbach Alpha Reliability Test and Nonparametric Kruskal-Wallis Test by using SPSS (Statistical Package for Social Sciences) 13.0 for Windows.

The findings of the descriptive statistics showed that students expect the instructors to attach more importance to the activities improving their speaking, listening and writing skills. Furthermore, the results displayed that nearly 73 percent of the instructors prefer the exams to be prepared by a testing office while more than half of the students expect them to be prepared by the instructor of the course. The results also revealed that both the students and the instructors believed it was necessary to use other test techniques in addition to the multiple-choice test technique commonly used in the exams.

According to the results of the One-way ANOVA, the more successful the students are, the more satisfied they are with the exams' different characteristics. As for the instructors, Nonparametric Kruskal-Wallis Test results indicated that there occurred no significant differences between instructors' educational background and the objective test techniques that they use in their classrooms. Additionally, it was found out there were no significant differences between instructors' educational background and their ideas on the objective test techniques that can be used in the exams. However, the more experienced the instructors are, the more efficient they find the exams prepared by the testing office.

Another important finding was that although their order of preferring objective test techniques slightly differs, the first eight test techniques that the students and instructors preferred in the exams were completely same.

The study concludes that both the students and the instructors have some doubts about the efficiency of the testing office's current practices. Therefore, for more efficient exams, test constructors can include the eight objective test techniques [(1) multiple-choice questions, (2) matching, (3) ordering tasks, (4) completion, (5) true-false questions, (6) short-answer questions, (7) error correction and (8) word changing], which were commonly preferred by the instructors and the students, into the Compulsory English Language Course Exams. In addition to the centrally administered achievement tests of this course, instructors should use teacher-made achievement tests and take the scores that students get from these tests into consideration while assessing their learners' language performance. Moreover, having a testing office with test constructors specialized just at testing will be a good idea for preparing better and more efficient tests.

## ÖZET

“Öğrenci ve Öğretmenlerin Üniversite Seviyesinde Dil Becerilerini Değerlendirmede Kullanılan Test Tekniklerine İlişkin Algılamaları” adlı bu çalışma, öğrenci ve okutmanların Çanakkale Onsekiz Mart Üniversitesinde dil performansını ölçmede kullanılan Zorunlu İngilizce Dersi Sınavlarına ilişkin algılamalarını tespit etmeyi amaçlamaktadır. Ayrıca; çalışma öğrenci ve okutman görüşlerini dikkate alarak bu sınavlarda çoktan seçmeli test tekniğinin yanı sıra başka hangi objektif test tekniklerinin kullanılabilceğini belirlemeyi hedeflemektedir.

Bu betimsel çalışmada nicel araştırma yöntemi kullanılmıştır. Yukarıda belirtilen hedefleri gerçekleştirmek amacıyla araştırmacı tarafından iki adet anket hazırlanmış ve anketler 367 öğrenciye, 33 okutmana uygulanmıştır. Anketteki maddelerin iç tutarlılık analizi yapıldıktan sonra, hem öğrenci anketi için hem de okutman anketi için kabul edilebilir Alpha güvenilirlik değerleri bulunmuştur. Araştırma için gerekli olan veri okutmanlardan ve öğrencilerden bu anketler aracılığıyla toplanmıştır. Okutman anketi Çanakkale Onsekiz Mart Üniversitesinde zorunlu İngilizce dersi okutmanı olarak görev yapmış ya da yapıyor olan tüm okutmanlara uygulanmıştır. Çalışmaya katılan öğrencilerin hepsi ikinci sınıf öğrencileridir ve Zorunlu İngilizce Dersini bu çalışma uygulanmadan bir yıl önce almışlardır.

Anketler aracılığıyla elde edilen veriler, tanımlayıcı istatistikler (Descriptive Statistics), tek faktörlü varyans analizi (One-way ANOVA), bağımsız örneklem T-Testi (Independent Samples T-Test), Cronbach Alpha güvenilirlik testi (Cronbach Alpha Reliability Test) ve parametrik olmayan Kruskal-Wallis Testi (Nonparametric Kruskal-Wallis Test) kullanılarak SPSS 13.0 istatistik programı ile analiz edilmiştir.

Tanımlayıcı istatistik sonuçlarına göre; öğrenciler okutmanlardan konuşma, dinleme ve yazma becerilerini geliştiren aktivitelere daha çok önem vermelerini beklemektedirler. Aynı zamanda, sonuçlar yaklaşık olarak okutmanların yüzde 73’ünün sınavların ölçme değerlendirme birimi tarafından hazırlanmasını tercih etmelerine karşın, öğrencilerin yarısından fazlasının ise sınavların sorumlu öğretim elemanı tarafından hazırlanmasını tercih ettiklerini göstermektedir. Sonuçlar, öğrencilerin ve okutmanların sınavlarda yaygın olarak kullanılan çoktan seçmeli test tekniğinin yanı sıra diğer test tekniklerini kullanmanın da gerekli olduğunu düşündüklerini ortaya koymuştur.

Tek faktörlü varyans analizi sonuçları; öğrencilerin başarıları arttıkça, onların sınavın farklı özelliklerinden daha çok memnun olduklarını gösterir. Okutmanlara gelince; parametrik olmayan Kruskal-Wallis testi, okutmanların eğitim düzeyleriyle sınıflarda kullandıkları test teknikleri ve yine eğitim düzeyleriyle sınavlarda kullanılabilceğini düşündükleri test teknikleri arasında anlamlı bir fark bulunmadığını göstermiştir. Ancak, okutmanların deneyimi arttıkça, ölçme değerlendirme birimi tarafından hazırlanan sınavları daha etkin buldukları ortaya çıkmıştır.

Çalışmadaki bir diğer önemli sonuç ise öğrencilerin ve okutmanların test tekniklerini tercih etmedeki sıralamalarında ufak farklılıklar olmasına karşın, her iki grubunda tercih ettiği ilk sekiz objektif test tekniğinin tamamıyla aynı olmasıdır.

Bu çalışmanın sonunda hem öğrencilerin hem de okutmanların ölçme değerlendirme biriminin şu anki çalışmalarının etkinliğine ilişkin bazı kuşkuları olduğu sonucuna ulaşılmıştır. Bu yüzden, daha etkin sınavlar için, sınav hazırlayanlar okutman ve öğrencilerin ortak olarak tercih ettikleri sekiz objektif test tekniğini [(1) Çoktan Seçmeli Sorular, (2) Karşılaştırma, (3) Sıraya Koyma, (4) Boşluk Doldurma, (5) Doğru-Yanlış Soruları, (6) Kısa Cevaplı Sorular, (7) Hata Düzeltme, (8) Kelime Değiştirme] Zorunlu İngilizce Dersinin sınavlarında kullanabilirler. Merkezi olarak yürütülen başarı sınavının yanı sıra; okutmanlar bu derste kendi hazırladıkları başarı testlerini de kullanmalı, öğrencilerin bu testlerden alacakları notları değerlendirme sürecinde dikkate almaları gerekmektedir. Ayrıca, sadece test etme üzerine uzmanlaşmış kişilerden oluşan ölçme değerlendirme birimi daha iyi ve daha yetkin sınavlar hazırlamada faydalı olacaktır.

## ACKNOWLEDGEMENTS

I would like to express my gratitude to many people who helped me complete this thesis. I owe special thanks to my supervisor Prof. Dr. Dinçay KÖKSAL who made this study possible. Without his guidance, encouraging criticism and support at every stage of the study, this thesis would never come true.

Special thanks must also go to Asst. Prof. Dr. İsmail Hakkı ERTEN, Asst. Prof. Dr. Ece ZEHİR TOPKAYA, Asst. Prof. Dr. Aysun YAVUZ and Asst. Prof. Dr. Cevdet YILMAZ, for the courage they gave me to complete the study and for the guidance they provided whenever I needed.

I also need to thank to Assoc. Prof. Dr. Ahmet AYPAY and Research Assistant Durmuş ÖZBAŞI for their great support at the construction process of the questionnaires and helping me with initial SPSS designing and the data analysis.

I would also like to thank my colleagues and to the second year students who have contributed to the development of this study by devoting their time on the questions in the questionnaires.

Finally, I'm deeply grateful to my wife and my family. Thank you for your patience and moral support at every stage of my study.

*“Behind every successful man, there is a strong woman.”*

*- Anonymous –*

*I have the strongest one.*

**TO MY WIFE...**



## TABLE OF CONTENTS

Title	
Abstract.....	i
Özet.....	iii
Acknowledgements.....	v
Dedication.....	vi
Table of Contents.....	vii
List of Tables.....	xi

### CHAPTER ONE

#### INTRODUCTION

1.0 Introduction.....	1
1.1 Background of the Study.....	1
1.2 Purpose of the Study and Research Questions.....	5
1.3 Significance of the Study.....	6
1.4 Assumptions of the Study.....	7
1.5 Limitations of the Study.....	7
1.6 Definitions.....	8
1.7 Scope of the Study.....	9
1.8 Summary.....	9

### CHAPTER TWO

#### LITERATURE REVIEW

2.0 Introduction.....	10
2.1 Testing.....	10
2.1.1 Evaluation and Measurement.....	12
2.1.2 Language Teaching and Testing.....	13
2.2 Basic Considerations in Language Testing.....	15
2.2.1 Reliability.....	15
2.2.2 Validity.....	16
2.2.2.1 Content Validity.....	17
2.2.2.2 Construct Validity.....	18
2.2.2.3 Criterion-related Validity .....	19
2.2.2.4 Face validity.....	20

2.2.3	Washback.....	21
2.2.4	Discrimination.....	21
2.2.5	Authenticity.....	22
2.2.6	Interactiveness.....	23
2.2.7	Administration.....	23
2.3	Types of Testing.....	24
2.3.1	Aptitude Tests.....	25
2.3.2	Placement Tests.....	26
2.3.3	Diagnostic Tests.....	26
2.3.4	Proficiency Tests.....	27
2.3.5	Achievement Tests.....	27
2.3.5.1	Teacher-made (Non-standardized) Achievement Tests.....	28
2.3.5.2	Standardized Achievement Tests.....	29
2.4	Test Techniques Used to Assess Language Performance at University Level: Students' and Teachers' Perceptions.....	31
2.4.1	Objective Tests for Standardized Achievement Tests.....	34
2.4.1.1	Multiple-choice Questions.....	36
2.4.1.2	Short-answer Questions.....	39
2.4.1.3	True-False Questions.....	41
2.4.1.4	Matching.....	42
2.4.1.5	Completion.....	43
2.4.1.6	Cloze Test.....	45
2.4.1.7	C-test.....	46
2.4.1.8	Cloze Elide Tests.....	46
2.4.1.9	Ordering Tasks (Rearrangement) .....	47
2.4.1.10	Error Correction.....	49
2.4.1.11	Transformation.....	50
2.4.1.12	Combination and Addition.....	50
2.4.1.13	Word Changing.....	51
2.4.2	Subjective Tests as Objective Ones.....	51
2.5	Summary.....	52

## **CHAPTER THREE**

### **METHODOLOGY**

<b>3.0</b> Introduction.....	53
<b>3.1</b> Research Method Used in the Study.....	53
<b>3.1.1</b> Construction of the Questionnaires.....	53
<b>3.1.2</b> Description of the Questionnaires.....	55
<b>3.2</b> Administering the Questionnaires.....	60
<b>3.2.1</b> The Pilot Study.....	60
<b>3.2.1.1</b> Setting.....	60
<b>3.2.1.2</b> Participants.....	60
<b>3.2.1.3</b> Materials.....	61
<b>3.2.1.4</b> Data Collection.....	61
<b>3.2.1.5</b> Data Analysis.....	62
<b>3.2.1.6</b> Implications for the Main Study.....	63
<b>3.2.2</b> Main Study.....	63
<b>3.2.2.1</b> Setting.....	63
<b>3.2.2.2</b> Sampling and the Participants.....	64
<b>3.2.2.3</b> Materials.....	68
<b>3.2.2.4</b> Data Collection.....	68
<b>3.2.2.5</b> Data Analysis.....	69
<b>3.3</b> Summary.....	69

## **CHAPTER FOUR**

### **FINDINGS AND DISCUSSION**

<b>4.0</b> Introduction.....	70
<b>4.1</b> Results and Discussion.....	70
<b>4.2</b> Summary.....	87

**CHAPTER FIVE**  
**CONCLUSIONS, SUGGESTIONS AND IMPLICATIONS**

<b>5.0</b> Introduction.....	88
<b>5.1</b> Summary of the Study.....	88
<b>5.2</b> Conclusions.....	89
<b>5.3</b> Suggestions.....	91
<b>5.3.1</b> Suggestions for the Instructors.....	91
<b>5.3.2</b> Suggestions for the Test Constructors.....	92
<b>5.4</b> Implications for Further Study.....	93
<b>REFERENCES</b> .....	95

**APPENDICES**

**Appendix A** : Open-ended Questions to Construct the Third Part of the Students’  
Questionnaire

**Appendix B** : Open-ended Questions to Construct the Third Part of the Instructors’  
Questionnaire

**Appendix C** : Piloting the Questionnaire

**Appendix D** : Students’ Questionnaire for the Main Study

**Appendix E** : Instructors’ Questionnaire for the Main Study

## LIST OF TABLES

<b>Table 1:</b> Third Part of the Students' Questionnaire.....	57
<b>Table 2:</b> Third Part of the Instructors' Questionnaire.....	58
<b>Table 3:</b> Items of the Fourth Part of the Instructors' Questionnaire.....	59
<b>Table 4:</b> Internal Consistency Reliability (Cronbach Alpha Coefficient) for the Fourth Parts of the Questionnaires in Piloting.....	62
<b>Table 5:</b> Gender Distribution of the Instructors Involving in the Study.....	64
<b>Table 6:</b> Experience of the Instructors Involving in the Study.....	65
<b>Table 7:</b> Educational Background of the Instructors Involving in the Study.....	65
<b>Table 8:</b> Instructors' BA Degree Graduation Departments.....	65
<b>Table 9:</b> Instructors' MA Degree Graduation Departments.....	66
<b>Table 10:</b> Instructors' Background Knowledge on Testing and Evaluation.....	66
<b>Table 11:</b> Stratified Random Sampling for Students.....	67
<b>Table 12:</b> Internal Consistency Reliability (Cronbach Alpha Coefficient) for the Fourth Parts of the Questionnaires in the Main Study.....	69
<b>Table 13:</b> Results of the Kruskal-Wallis Test Regarding Difference between the Activities that are used by the <b>Instructors</b> , and the Activities that the <b>Students</b> Expect to be used to Develop Language Skills and Areas.....	70
<b>Table 14:</b> Instructors' Ideas on Whom to Organize the Exams.....	72
<b>Table 15:</b> Students' Ideas on Whom to Organize the Exams.....	72
<b>Table 16:</b> The Instructors' Reasons for Preferring the Testing Office.....	73
<b>Table 17:</b> The Instructors' Reasons for Preferring the Instructor of the Course.....	73
<b>Table 18:</b> The Students' Reasons for Preferring the Testing Office to Organize the Exams.....	74
<b>Table 19:</b> The Students' Reasons for Preferring the Instructor of the Course to Organize the Exams.....	75
<b>Table 20:</b> The Instructors' Perceptions of the Exams.....	75
<b>Table 21:</b> The Students' Perceptions of the Exams.....	76
<b>Table 22:</b> Results of ANOVA Regarding the Difference between Students' Success and their Perceptions of the Exams.....	77
<b>Table 23:</b> Mean Values of the Unsuccessful, Successful and Very Successful Students' Ideas.....	79

<b>Table 24:</b> The Results of Kruskal-Wallis Test Regarding the Difference between Instructors' Job Experience and their Perceptions of the Exams.....	80
<b>Table 25:</b> The Results of Independent Samples T-Test Regarding the Difference Between Students' and Instructors' Ideas on How Frequently Objective Test Techniques Are Used in the Class.....	82
<b>Table 26:</b> Students' and Instructors' Perceptions of the Objective Test Techniques that can be Used in the Standardized Exams.....	84
<b>Table 27:</b> <i>The Results of</i> Kruskal-Wallis Test Regarding the Difference between Instructors' Educational Backgrounds and the Objective Test Techniques They Use.....	85
<b>Table 28:</b> <i>The Results of</i> Kruskal-Wallis Test Regarding the Difference between Instructors' Educational Backgrounds and Their Ideas on the Objective Test Techniques that Can be Used in the Exams.....	86

## **CHAPTER I**

### **INTRODUCTION**

#### **1.0 INTRODUCTION**

Teaching has been very important for ages. Nearly all types of teaching need testing at the end. In this chapter, it is aimed to present the reasons why testing is so important in the teaching-learning process and why the researcher decided to work on present study which is about students' and instructors' perceptions of the test techniques used in the standardized achievement test of the Compulsory English Language Course at Çanakkale Onsekiz Mart University (hereafter COMU)". The importance of testing and the things to be taken into consideration while testing are given in 1.1 entitled *Background of the study*, and the aim of the study and the research questions are given in 1.2 as *Purpose of the Study and Research Questions*. Later, in 1.3 the *Significance* of the study is highlighted. In 1.4 the *Assumptions*, and in 1.5 the *Limitations of the study* are presented. Finally, in 1.6 the *Scope of the Study* is introduced before dealing with the literature review.

#### **1.1 BACKGROUND OF THE STUDY**

Teaching and learning is a broad process. The aim of teaching foreign language is to provide students with the necessary language skills (speaking, listening, reading, writing) and areas (grammar, vocabulary, pronunciation). Firstly, teaching process necessitates a well-prepared plan in order to reach instructional objectives and it does not finish when the students have learned the subject. Then, the students' success should be evaluated and measured at the end of this process.

In this respect, evaluation and measurement become indispensable part of the teaching and learning process as testing helps teaching reach its aims. To test accurately, the teacher should firstly know what a test is and what it consists of.

Many definitions of a test have been provided by several authors. However, H.D. Brown's (2001:384-385) definition of a test is well-founded. He defines the test as "a method of measuring a person's ability or knowledge in a given domain". Firstly, it is a "*method*" which consists of different techniques, procedures and items. Secondly, "*measuring*" is the aim of each test. Also, testing measures "*a person's*" ability or knowledge. Thus, test takers' characteristics are really important in testing. Furthermore, students' "*ability*" is tested in a given "*domain*".

Testing is really important for several reasons. To Madsen (1983:3), "testing is an important part of every teaching and learning experience". In his opinion (1983), testing helps both the students and the teacher of a class. It helps teachers find an answer to the question of whether they have been effective in their teaching or not. In addition to diagnosing students' efforts, testing diagnose the teachers' own efforts. Testing answers all the questions to be answered about our own way of teaching. It gives some ideas to the teachers for the future evaluation. Testing also helps the teacher to make decisions which will help his or her teaching. Popham (2003) summarizes these decisions as:

- decisions about the nature and purpose of the curriculum,
- decisions about students' prior knowledge,
- decisions about how long to teach something,
- decisions about the effectiveness of the instruction.

When the teacher makes such decisions after testing, he/she can enhance his or her way of teaching.

As for students, tests help them (Madsen, 1983:3-4):

- ... create positive attitudes toward [their] class,
- ... learn the language by their diagnostic characteristics,
- ... create positive attitudes toward instruction by giving students a sense of accomplishment...,
- ... learn the language by requiring them to study hard, emphasizing course objectives, and showing them where they need to improve.



Working together with teaching, testing has much more benefits for both the learners and the teachers. As for the learners, testing (1) promotes meaningful involvement of the learners with the material, (2) gives chance to review the material covered in the course, (3) and provides learners with feedback about their language performance. When it comes to the teachers, testing (1) helps them determine the objectives of the instruction, (2) provides them feedback for improving future assessment, and (3) shows their students' strengths and weaknesses (Cohen, 1994). When they work hand in hand successfully, teaching and testing reach their aims. Thus, students learn effectively.

As suggested above, tests affect teaching process. In order to affect teaching process positively, teachers should take some important points into consideration while constructing their tests. In order to be called as "a good test", a test should have some characteristics such as reliability, validity, washback, discrimination, authenticity, interactiveness, and practicality.

Cohen (in Celce-Murcia, 2001:516) claims that a good test can be used for twelve different purposes:

... five administrative purposes (achievement, placement, exemption, certification, promotion), four instructional purposes (diagnosis, evidence of progress, feedback to the respondent, evaluation of teaching or curriculum), and three research purposes (evaluation, experimentation, knowledge about language learning and language use).

However, a normal task cannot include all these purposes. To Cohen, the basic distinction is often between *proficiency tests* that are usually for administrative purposes and *achievement tests* used for the assessment of instructional outcomes.

Achievement tests measure what students have learned at the end of the teaching process. These tests are directly related to *the language courses*. As this study is about students' and instructors' perceptions of the Centrally Administered Achievement tests used at *Compulsory English Language Courses*, Standardized Achievement Tests (SATs) are the main focus of this research study. At COMU, all

the faculty students have “Compulsory English Language Courses” in their first year. At the end of the course, a SAT is used to assess students’ language performance.

Only multiple-choice test items are used to assess students’ language performance at the end of Compulsory English Language Courses. Multiple-choice question form used at COMU only tests the ability to recognize correct grammar and usage. However, they do not intend to test the ability to produce.

To Bachman and Palmer (1996), testing is a means to evaluate the educational programme and to give feedback on teaching. Thus, to get an accurate feedback on teaching, it is important to apply accurate achievement tests to the students of COMU. In the opinion of Hughes (1989), there are two main reasons for a test to be inaccurate. The first is about test content and techniques. For example, if the writing skill is only tested by multiple choice items, the students practise such items rather than the skill of writing. Thus, the test becomes inaccurate. This is the case at COMU as the coursebook aims to include all language skills; however, students are tested through only multiple-choice test technique in their exams. The second reason is the lack of reliability. To him, unreliability has two origins: “features of the test itself, and the way it is scored” (p.3). The way the teachers score the tests at COMU Compulsory English Language Courses can be totally same and reliable; however, features of the test and the technique used in these tests may not be accurate.

All in all, in the light of the studies and the facts mentioned above, it is really important to prepare an achievement test which tests what is intended to test. In order to prepare accurate tests, it is crucial to have an idea of students’ and instructors’ perceptions of the test techniques that can be used in standardized achievement tests of COMU. Their perceptions of different test techniques to be used to assess students’ language performances will be of great importance to the future practices of the test constructors in the “Testing Office”. What students and instructors think about the present studies of the test constructors and what they suggest for the future studies will be beneficial for preparing accurate tests.

## 1.2 PURPOSE OF THE STUDY AND RESEARCH QUESTIONS

The main purpose of this study is to have an idea on the students' and the instructors' perceptions of the Compulsory English Language Course exams used to assess language performance at COMU. Moreover, this study aims to determine what other objective test techniques can also be used in these exams in addition to the multiple-choice test technique while taking all the students' and the instructors' opinions into consideration.

The study addresses the following research questions:

**RQ1** Are there any differences between the activities that are used by the instructors and the activities that the students expect to be used in their classrooms in order to develop different language skills and areas?

- Which language skill or area do the students expect to be developed most in their lessons?
- Which language skill or area do the instructors try to develop most in their lessons?

**RQ2** Do the students and the instructors prefer the Compulsory English Language Course exams to be prepared by the testing office or by the instructor of the course?

- What is their most common reason for preferring either testing office or the instructor for the preparation and organization process of the exam?

**RQ3** What do the instructors and the students think about the exams that have been prepared by the test constructors of the testing office so far?

- Is there a significant difference among students' thoughts about the exams prepared by the testing office in terms of their success?
- Is there a significant difference among instructors' thoughts about the exams prepared by the testing office in terms of their job experience?

**RQ4** What are students' and instructors' perceptions of the objective test techniques that are used and that can be used in the Standardized Achievement Tests of Compulsory English Language Course?

- Are there any significant differences between the students' and instructors' ideas on how frequently objective test techniques are used in the class?
- Which objective test techniques do the instructors use most and least frequently in the classroom activities of Compulsory English Language Courses?
- What other objective test techniques can be used in the mid-term and final exams of the course in the future?
- Do the objective test techniques that are used by the instructors in their classroom activities differ in terms of instructors' educational backgrounds?
- Do the instructors' ideas on which objective test techniques can be used in Centrally Administered Achievement tests differ in terms of their educational backgrounds?

### **1.3 SIGNIFICANCE OF THE STUDY**

This study will contribute to the literature on students' and instructors' perceptions of the test techniques used to assess students' language performance at university level. There have always been arguments about the way language is being tested at the meetings of the instructors. The way language is being tested has been criticized by the instructors; however, their specific ideas on how to test students' language performance better have never been questioned up to now. This study is intended to be useful for the instructors of English at COMU as the findings of the study will reflect the instructors' ideas on the current practices of the testing office. Instructors' ideas on its current practices will help the testers of the testing office prepare better exams. Furthermore; not only the instructors, but also the students are the part of the teaching and learning process. Thus, asking students' ideas on the ways they are being tested is as crucial as asking that of instructors.

As it attempts to demonstrate the students' and instructors' ideas on the appropriateness of the current testing system, this study will provide some suggestions for constructing and administering better language tests. All in all, what makes this study important is that it presents the students' and instructors' perceptions of what has been done so far, what is being done now, and what can be done in the future to test students' language performance at COMU.

#### **1.4 ASSUMPTIONS OF THE STUDY**

The main assumptions of the study are as follows:

1. The ideas of the instructors and the experts (at ELT) are enough to prepare and apply the questionnaire which is used for data collection.
2. Students chosen for this research are all in their second years at the university. They all had the "Compulsory English Language Course" the year before the study was conducted. Therefore, they are assumed to answer the questions in the questionnaire without any fear of their instructors or the possibility of the changes in the testing system.
3. The researcher made a *Stratified Random Sampling* of the whole students' population. It is assumed that the students participating in the study represent the total number of the students.
4. The findings of the study will reflect the real facts about the students' and instructors' perceptions of the testing system at COMU.
5. All participants of the study are assumed to have taken part willingly and to have given answers with complete frankness.

#### **1.5 LIMITATIONS OF THE STUDY**

This research is limited to the opinions of 33 instructors and 367 students at COMU. The size of the sample is, for the instructors, nearly what it has to be. Thus, it is satisfactory. However, for the students, it is the major limitation of the study. Stratified Random Sampling of the students was only done considering the number

of the students in different faculties (See Table 11). Stratification was not done in terms of the departments as it would be really difficult to find samples from all departments. Thus, volunteer and available second-year students were chosen for the study according to the number of the students in different faculties. Furthermore, the students involving in the main study had taken the midterm and final exams of the Compulsory English Language Course nearly ten months before the main study was carried out. Therefore, memory effect can be another limitation for this study.

This study is also limited with the questions asked in the questionnaire. It may not reflect all other opinions of each student or that of each instructor. Using only a questionnaire for data collection can be considered as another limitation. However, the number of the students to be questioned, the difficulty of reaching the instructors as they work in different faculties or colleges and the limited time to carry out this research forced the researcher to use a practical method for data collection, which is *questionnaire*.

## 1.6 DEFINITIONS

In this part of the thesis, some important terms that can be unfamiliar or unknown to the readers are explained and defined.

**Centrally Administered Achievement Test:** Any assessment device that is administered and scored in a standard, predetermined manner.

**Compulsory English Language Course:** A compulsory course that students have to take generally in their first years. This course aims to take students from false beginner to A1 level.

**Testing Office:** A group of instructors who prepare and organize the mid-term and final exams of the Compulsory English Language Courses. This group is assigned by the head of the Foreign Languages Department. The instructors working for the testing office are also responsible for teaching Compulsory English Language Courses.

## **1.7 SCOPE OF THE STUDY**

This thesis consists of five chapters. In Chapter One, the background, the purpose, the research questions, the significance, the assumptions, and the limitations of the study are presented. Chapter Two presents a review of the literature on testing and especially on the objective test techniques that are used in standardized achievement tests. Various studies and their findings are discussed in this chapter. Chapter Three is about the methodology used in the research. The research method used in the study is stated in this chapter. The construction process of the questionnaires is described. Moreover; the setting, participants, materials, and the data collection and analysis processes of the pilot study and the main study are explained in detail. Chapter Four presents the findings of the study and the discussions about those findings. Finally; a conclusion, in which a brief summary of the study is provided, suggestions for the instructors and the test constructors, and implications for further study are available in Chapter Five.

## **1.8 SUMMARY**

Some basic literature on the importance of testing both for the students and for the teachers was presented in this chapter. Then, the purpose of the study and the research questions were introduced. Later on, the significance, assumptions, and limitations of the study were explained in detail in different sections. Finally, what parts the thesis consists of was presented in the previous section, scope of the study.

## **CHAPTER II**

### **LITERATURE REVIEW**

#### **2.0 INTRODUCTION**

Different studies related to testing are reviewed in this chapter. The main aim of this chapter is to state the main issues about testing. Then, to review some studies on students' and teachers' perceptions of the test techniques used to assess language performance at university level.

In this chapter; firstly, the terms: testing, evaluation, measurement, language teaching and language testing are explained by means of different studies and definitions which have been provided by different researchers. Then, some basic considerations in language testing to create an effective test are given in detail. These are also known as the characteristics of a good test: reliability, validity (content, construct, criterion-related, face validity), washback, discrimination, authenticity, interactiveness and administration. After stating the basic considerations in language testing, types of testing (aptitude, placement, diagnostic, proficiency, and achievement tests) are mentioned.

Having explained all these issues, the researcher provides the literature on the main issue of the thesis: the students' and the instructors' perceptions of the test techniques used at standardized achievement tests. Finally in this part, the researcher aims to state the advantages and disadvantages of test techniques that can be used at standardized achievement tests.

#### **2.1 TESTING**

Teaching a foreign language involves providing students with basic language skills (speaking, listening, writing and reading) and language areas (grammar,



vocabulary and pronunciation). However, teaching does not finish when students have learned the subject matter. Testing can help teaching reach its aims. As Madsen (1983:3) suggests “testing is an important part of every teaching and learning experience”.

For accurate testing, an effective teacher of English should firstly know what a *test* is and what its constituents are. So far so many definitions of a *test* have been provided. To start with the dictionary definition of a *test*, it is “a set of questions to measure someone’s knowledge or ability” (Cambridge Learner’s Dictionary, 2004: 669). This definition is a general one. However, the reader can get the broad understanding of the meaning. H. D. Brown (2001:384-385) defines the term *test* as “a method of measuring a person’s ability or knowledge in a given domain”. So, it is first a “*method*”. That is, it mainly consists of different techniques, procedures, and items. Next, it has the aim of “*measuring*”. Also, it measures “*a person’s*” ability or knowledge. Hence, test takers characteristics are really important in testing. Then, what is being tested in a test is “*ability*”. Finally, test measures this ability in a given “*domain*”. Furthermore, Bachman (1990:20) provides another definition as follows:

A test is a measurement instrument designed to elicit a specific sample of an individual’s behavior. As one type of measurement, a test necessarily quantifies characteristics of individuals according to explicit procedures. What distinguishes a test from other types of measurement is that it is designed to obtain a specific sample of behaviour.

Also Genesee and Upshur (1996:141) express their ideas on what the test constituents are. In their opinion:

A test is, first of all, about something. That is, it is about intelligence or European history, or second language proficiency. In educational terms, tests have subject matter or content. Second, a test may consist of only one task, such as writing a composition, or a set of tasks, such as in a lengthy multiple-choice examination in which each question can be thought of as a separate task. ... Third, tests yield scores that represent attributes or characteristics of individuals. ... Test scores along with the frame of reference used to interpret them is referred to as measurement. Thus, tests are a form of measurement.

As it can be understood from the definitions above, there is always something to be measured and evaluated in teaching and learning process. Hence, the distinction between *evaluation* and *measurement* is clarified below before the explanation of the relationship between language teaching and language testing.

### 2.1.1 EVALUATION AND MEASUREMENT

The terms evaluation and measurement are usually used as synonyms; however, they are definitely different processes. Evaluation, which is more comprehensive of the two terms, also comprises measurement. To Green (1970:5), “actually the two are, in a sense, the inseparable sides of a coin which are used in a working partnership”. Green’s argument about these terms is well-founded. Teachers of English should effectively use both evaluation and measurement in teaching process.

In their article about evaluation, Dickins and Germanie (1995) claim that evaluation involves not only the testing of students’ knowledge but also the teacher’s development and the parts of the curriculum. Thus, measurement which is done in order to see learners’ performance is only a part of evaluation.

In his paper on *Key Concepts in ELT*, Murphy (2000:210) explains the purpose of evaluation as:

... to determine the extent to which a programme or intervention is worthwhile, and to aid decision making through the purposeful gathering of information which is analysed and reported to stakeholder-interested parties who have a ‘stake’ in the activity evaluated.

As it can be understood from the purpose of it, evaluation is primarily about decision making. Tests help the teacher collect the necessary information about their students. This information helps the teacher make decisions about their students’ knowledge and ability. Genesee and Upshur (1996:3) state that:

Language evaluation involves many different kinds of decisions: decisions about placement of individual students in particular streams, levels or courses of instruction; about ongoing instruction; about planning new units of instruction and revising units that have been used before; about textbooks or other materials; about student homework; about instructional objectives and plans; and about other aspects of teaching and learning.

Thus, evaluation becomes really important part of the teaching and learning process. To evaluate students’ performance, the teacher needs to measure their knowledge and ability either by using oral or written form of testing. While dealing with measurement, the teacher should attach importance to some points, which will be explained in detail in part 2.2 *Basic Considerations in Language Testing*. However, it

is really important to highlight the relationship between *Language Teaching and Testing* beforehand. Hence, the following section will be about this relationship and its importance.

### **2.1.2 LANGUAGE TEACHING AND TESTING**

For years, there has been a strong relationship between language teaching and testing. As Popham (2003) suggests, when teachers do their instructional jobs well, their students will be successful in their tests. He also claims that the way the teacher tests can influence how well he or she teaches. It is not enough to accept the idea that testing can help teaching. Teachers should also put this idea into practice.

As also discussed in the previous part, evaluation is primarily about decision making. Testing helps the teacher to make decisions which will help his or her teaching. Popham (2003) summarizes these decisions as:

- decisions about the nature and purpose of the curriculum,
- decisions about students' prior knowledge,
- decisions about how long to teach something,
- decisions about the effectiveness of the instruction.

Having made such decisions after testing the teacher can enhance his or her way of teaching.

Language testing helps teaching in many ways. In the opinion of Davies (1990:1, quoted in Bostan, 2005:8) "language testing provides goals for language teaching and it monitors both teachers' and learners' success in reaching these goals". Bachman (1990:54) points out another use of testing as "to provide information for making decisions, that is, for evaluation". Tests have many more uses. Salkind (2006:12) states that they are used for "selection, placement, diagnosis, hypothesis testing, and classification". Wherever, whenever and however teachers teach, they need these uses of testing to reach their instructional goals.

Furthermore, Heaton (1988:5) makes it clear that “both testing and teaching are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the other”. Working together with teaching, testing has many benefits both for the learners and for the teachers. As for the learners, testing (1) promotes meaningful involvement of the learners with the material, (2) gives chance to review the material covered in the course, (3) and provides learners with feedback about their language performance. When it comes to the teachers, testing (1) helps them determine the objectives of the instruction, (2) provides them feedback for improving future assessment, and (3) shows their students’ strengths and weaknesses (Cohen, 1994).

Similarly, Bachman (1990) attaches great importance to testing. He (1990:22) believes that “... the value of tests lies in their capability for eliciting the specific kinds of behavior that the test user can interpret as evidence of the attributes or abilities which are of interest”. Moreover, a good test should have a “positive effect on learning and teaching and should generally result in improved learning habits” (Heaton, 1988:5). This means that when they work hand in hand successfully, teaching and testing reach their aims. Thus, students’ effective learning occurs.

Like many other authors, Madsen (1983) also makes it clear that testing helps students and teachers in many ways. To him (1983:3-4) tests help students:

- ... create positive attitudes toward [their] class,
- ... learn the language by their diagnostic characteristics,
- ... create positive attitudes toward instruction by giving students a sense of accomplishment...,
- ... learn the language by requiring them to study hard, emphasizing course objectives, and showing them where they need to improve.

Madsen (1983) also discusses how testing helps teachers of English. It helps teachers find an answer to the question of whether they have been effective in their teaching or not. In addition to diagnosing students’ efforts, testing diagnose the teachers’ own efforts. Testing answers all the questions to be answered about our own way of teaching. It gives some ideas to the teachers for the future evaluation. All in all,

“...good tests can sustain or enhance class morale and aid learning.” (Madsen, 1983:5).

As it is clearly understood, good tests affects language teaching in a positive way. Following part provides a general view of a good language tests, their characteristics and basic considerations in language testing.

## **2.2 BASIC CONSIDERATIONS IN LANGUAGE TESTING**

As suggested above, tests affect teaching process. Both the teacher and the students become aware of their performances in teaching and learning by means of tests (Bostan, 2005). There are some important points to be taken into consideration while constructing tests. In order to be called as “a good test”, a test should have some characteristics such as reliability, validity (content, construct, criterion-related and face validity), washback, discrimination, authenticity, interactiveness, and administration (practicality).

The first two of the characteristics – reliability and validity are “essential measurement qualities” (Bachman & Palmer, 1996:19). However, others have the other components of a learning programme (Bachman & Palmer, 1996). Namely, while reliability and validity are mostly associated with measurement, the others have much to do with the other components of the language programme like the curriculum, school administration, teachers, students, coursebook, and etc.

All these basic considerations in language testing will be explained in detail in the following sections.

### **2.2.1 RELIABILITY**

To be reliable, “a good test should give consistent results” (Harmer, 2001:322). Thus, reliability can be defined as “consistency of measurement”

(Bachman & Palmer, 1996:19). That is to say, if a test is reliable, it gives same or approximate results when it is administered to different students at the same time or to the same student on different occasions (Bachman & Palmer, 1996; Harmer, 2001; Hughes, 1989; Weir, 1990).

Reliability is closely associated with error. Demirel (2002:212) defines reliability as “degree of purification from errors”. There are three types of error. Biased error and constant error are easier to detect, since their quantities and origins can be found in certain ways. On the contrary, the quantity and origin of chance error are not evident. If the teacher marks only one student’s exam paper wrong, it is chance error. Adding five points more for each student is a constant error. The teacher’s boosting hard-working students’ grades can be a good example of biased error (Demirel, 2002). The more the teacher makes errors, the more the test will be unreliable.

To Ahmann and Glock (1971), instrument-centered factors that affect reliability of a test are more damaging than student-centered factors. The assertion in this statement is, in my opinion, convincing. While student-centered factors can change from day to day and influence the student’s own grade, instrument-centered factors affect all students who have taken the test.

In brief, the teacher should be aware of his errors, instrument-centered factors, and student-centered factors. They should avoid all kinds of errors and try to use effective teaching and testing instruments to make their tests more reliable.

### **2.2.2 VALIDITY**

A test can be reliable; however, it can be totally invalid. A test’s validity can be explained as the extent to which the test measures what it is supposed to measure (H.D. Brown, 2001; J.D. Brown, 1996; Hughes, 1989; Salkind, 2006; Weir, 1990). To be valid, a test must measure a representative “sample behavior” (Gronlund,

1971:97). In other words, we should measure only one skill at a time. For example, a writing test item such as ‘Is overhead projector a practical device? Discuss.’ is invalid, because it requires both the knowledge of writing and educational materials at the same time. Unless a student knows the use of overhead projector, he/she cannot write anything about it.

Validity and reliability are closely related. However, “validity is a separate but equally important issue” (J.D. Brown, 1996:231). J.D. Brown (1996) makes it clear that reliability and validity are different test qualities although they are related. To him (1996:231) “... reliability is a precondition for validity but not sufficient for the purposes of judging overall test quality”. Thus, validity must also be carefully studied although a test seems to be reliable.

Alderson, Clapham, and Wall (1995:171) argue that:

It is best to validate a test in as many ways as possible. In other words, the more different ‘types’ of validity that can be established, the better, and the more evidence that can be gathered for any type of validity, the better.

If a test lacks validity, then it will not be able to test what it claims to test. Therefore, nobody (students, teachers, administrators, and etc.) will be able to get benefit from the test.

To make the tests more valid, the teachers should be careful about four main types of validity: content, construct, criterion-referenced, and face validity. The following sections give brief information about these types of validity.

### **2.2.2.1 CONTENT VALIDITY**

Gronlund (1971:97) defines the term content validity as “the extent to which a test measures a consistent sample of behavior”. To illustrate, if the teacher prepares an English test for the ninth class, it should include all related subjects in this degree. So, to have content validity, a test’s content must be the representative of the subject area being covered. However, of course a test cannot deal with every subject covered

on a program. Then, such a test would be quite long. Ethically, the teacher must select sample items from the total content. These sample items should make us sure that, if students know the sample material, the teacher can say that they have learned the whole material. So, as an ethic of measurement good sampling is really important. Hughes (1989:23) believes that the content of tests is often determined by “what is *easy* to test rather than what is *important* to test”. This is a complete danger for a test to lose its content validity. To Hughes (1989:23), “the best safeguard against this [danger] is to write full test specifications and to ensure that the test content is a fair reflection of these”. This must be the case while preparing the SATs of Compulsory English Language Courses.

All in all, to achieve content validity, it is important for the tester to plan the course content, and what has been taught till the test. Namely, the teacher should plan a list of everything that the test will cover. Then, he/she needs to ensure that the things in that list are sampled in the most suitable way. In brief, if a test covers what it should cover, we can say that that test has content validity. As a common rule of content validity, it can be stated that *measure what you are supposed to measure, nothing else*.

#### **2.2.2.2 CONSTRUCT VALIDITY**

The other type of validity that teachers should be aware of in terms of language tests is construct validity. Davidson, Hudson, and Lynch (1985, cited in H.D. Brown, 2001:389) suggest a way to look at a construct validity of a test. This way is to:

... ask the question “Does this test actually tap into the theoretical construct as it has been defined?” “Proficiency” is a construct. “Communicative competence” is a construct. “Self-esteem” is a construct. Virtually every issue in language learning and teaching involves theoretical constructs.



A teacher should be satisfied with the test in that it is a sufficient definition of a construct. H.D. Brown (2001:389) perfectly exemplifies construct validity as follows:

Let's say you have been given a procedure for conducting an oral interview. The scoring analysis for the interview weights several factors into a final score: pronunciation, fluency, grammatical accuracy, vocabulary use, and sociolinguistic appropriateness. The justification for these five factors lies in a theoretical construct that claims those factors as major components of oral proficiency. So, on the other hand, if you were asked to conduct an oral proficiency interview that accounted only for pronunciation and grammar, you could be justifiably suspicious about the construct validity of such a test.

As it is clear from the example above, construct validity has a lot to do with the relationship between a test and a particular view of language testing. If a test measures the particular view of language that it aims to measure, only then it can be called as a test having high construct validity.

### **2.2.2.3 CRITERION-RELATED VALIDITY**

Another way of understanding that a test is valid is “to see how far results on the test agree with those provided by some independent and highly dependable assessment of the candidate’s ability. This independent assessment is thus the criterion measure against which the test is validated” (Hughes, 1989:23). Namely, if there is “a relationship between test scores and some criterion...” (Bachman, 1990:248), it can be said that the test has high level of criterion-related validity.

To Genesee and Upshur (1996:66), “criterion-related validity can be determined using statistical procedures that quantify the degree of agreement between one type of assessment and a criterion”. To make it more clear, the case in the Compulsory and Voluntary Preparatory Classes of Çanakkale Onsekiz Mart University can be given as an example. In these preparatory classes, a placement test is used at the very beginning of the term to classify students according to their language levels. As this test has been used many times and has given effective results for many years, it can be considered as a reliable one. Thus, it can be used as a

criterion to make our classroom tests more valid. If the scores of our classroom tests and the placement test (criterion) show little agreement after the statistical procedures, our classroom tests cannot be considered as valid, then. However, if there is a high level of agreement between the scores of these tests, then our classroom tests may be considered as valid.

There are two types of criterion-related validity: concurrent validity and predictive validity. Concurrent validity is established when the criterion and the test are given to the candidates simultaneously (Hughes, 1989). On the other hand, predictive validity is about “the degree to which a test can predict candidates’ future performance” (Hughes, 1989:23).

#### **2.2.2.4 FACE VALIDITY**

The last form of validity explained here is face validity. It refers to “whether the test looks as if it is measuring what it is supposed to measure” (Cohen, 1994:40). Face validity is concerned with whether the test is perceived as valid by the teacher and the students. It has to do with what the outsider thinks of the test. For the test to have face validity, someone - this can be a student, a teacher, or someone from outside - must look at the test and be convinced that the test giving valid information. That is to say, face validity “... is not validity in technical issue; it refers not to what the test actually measures, but to what it appears superficially to measure” (Anastasi, 1982:136, quoted in Weir, 1990:26).

If a test does not have face validity, students may not accept it. Such a test will most probably be “... irrelevant, inappropriate, silly or childish...” (Anastasi, 1982:136, quoted in Weir, 1990:26) to the students and this will affect their learning process negatively. Thus, the teacher should also be careful about the *face validity* although it seems to be the least important one. In this aspect, proofreading gains importance. Before applying tests, a piloting should be done if it is possible.

### 2.2.3 WASHBACK

Ethically, good tests should have effects both in the classroom and beyond the classroom, both on the individual and on the society and educational systems. Applying a test implies certain values, and goals; and has consequences. As Bachman (1990:279) believes “tests are not developed and used in a value-free psychometric test-tube; they are virtually always intended to serve the needs of an educational system or of society at large”. Thus, Bachman and Palmer (1996:30) highlight, whenever the teachers use tests, they should use them “in the context of specific values and goals”. Those tests should have “specific consequences for, or impact on both the individuals and the system involved” (Bachman & Palmer, 1996:30).

McNamara (2000:73) defines **washback** as “the effect of tests on teaching and learning”. And he also (2000:74) emphasizes that “ethical language testing practice should work to ensure positive washback from tests”. Ensuring positive washback depends on many factors such as “local conditions in classrooms, the established traditions of teaching, the immediate motivation of learners, and frequently unpredictable ways in which classroom interactions develop” (McNamara, 2000:73). Those factors not only have impact on the individual but also on the society and the educational system. Hence, washback is not simply the effect of testing on teaching and learning in the classroom atmosphere. It also has the potential for affecting not only individual but the society and the educational system as well. All in all, as an ethic of measurement, in order to ensure positive washback effect from the test, the teachers should take the factors above into consideration while preparing and applying their test.

### 2.2.4 DISCRIMINATION

To Heaton (1988:165), an important characteristic of an effective test is “its capacity to discriminate among different candidates and to reflect the differences in

the performances of the individuals in the groups”. For example, if almost all the students score 80 out of 100 on the test, this test clearly lacks the quality of discrimination. It will be insufficient to discriminate between various students.

Although it is an important quality, discrimination may not be used in the classroom. “The extent of the need to discriminate will vary depending on the purpose of the test...” (Heaton, 1988:165). For example, the teacher can only want to know how well his/her students have learnt the subject matter taught. Thus, he/she will not be interested in discriminating his/her students. However, while choosing three students out of twenty for a competition, he/she will need to prepare a test to discriminate relative abilities of the students.

### **2.2.5 AUTHENTICITY**

This term refers to the relationship between the use of the target language and the test. Bachman and Palmer (1996:23) claim that “it is this correspondence that is at the heart of authenticity”. If anyone prepares a test task characteristics of which correspond to the task of the *target language use* (TLU), Bachman and Palmer (1996:23) describe that kind of a test as “relatively authentic”. They (1996:23) define authenticity as “the degree of correspondence of the characteristics of a given language test task to the features of a TLU task”.

To Bachman and Palmer (1996:24), one important reason for considering authenticity as important is that it has “potential effect on test takers perceptions of the test and, hence, on their performance”. If the topical content of the test is related to the TLU task, test becomes authentic. And this relevance can help the test takers promote a positive response to the test task and can thus help them do their best in their tests.

As the designers of language tests, teachers should implicitly or explicitly consider authenticity in designing their language materials. To illustrate, while

developing a reading test, teachers must always choose a topic which they think the test takers may read outside of the testing situation. It must be done for the sake of authenticity. Authenticity will help the test designer and the test takers a lot as they both can relate the test task and the TLU task. As a result, both the test designer and the test taker will benefit from that kind of authenticity which helps them perform at their best.

### **2.2.6 INTERACTIVENESS**

Another important issue to be taken into consideration is interactiveness of a test. Bachman and Palmer (1996:25) define the term as "... the extent and type of involvement of the test takers' individual characteristics in accomplishing a test task."

The interactiveness of a test can sometimes be considered as authenticity; however, they are different terms and different test characteristics. Both of them have potential effect on students' perceptions of the test and, thus, on their success. However, while authenticity relates the test task to the target language use, interactiveness relates the test taker to the test task (Bachman & Palmer, 1996). To illustrate, a case at the preparatory classes of Çanakkale Onsekiz Mart University will be really understandable. At a speaking final exam, students are sometimes asked to talk about any subject they want. This type of exam lacks authenticity as a person will never be asked to talk about whatever he/she wants in real life. Nevertheless, this will be highly interactive as students are actively involved in the test task. Students' being knowledgeable about the subject and students' interests demonstrate high level of interactiveness of the test.

### **2.2.7 ADMINISTRATION**

Administration of a test is synonymous with practicality. A test should be practical in terms of preparation, marking, materials and equipment. It is an

undeniable fact that the teacher should prepare his/her plan before the lesson. It is the same case while testing. He/She must plan again before constructing the test for that lesson.

Tests require special arrangements as they are used in the class. While testing one individual, the rest of the class should not be ignored. Besides the arrangements, time limitations ought to be taken into consideration. Time given must be consistent with the difficulty of test items. Sometimes it is crucial to provide students with some materials and equipments. Accordingly, size of the class should be added to the test organization. Another vital point to mention is marking. An effective teacher of English, as an ethic of measurement, must prepare his/her answer key before test application in order not to be hesitant. In short, the preparation and application of the test ought to be economical and practical in all aspects. Bachman and Palmer (1996:36) define a practical test as a one “whose design, development and use do not require more resources than are available”. Thus, the teacher must be aware of the resources that are available in his/her hand and prepare his/her tests accordingly.

It is an undeniable fact that if a test lacks the characteristic of practicality, it is nothing more than a waste of time even if it is valid and reliable. Therefore, this quality of a language test is as important as the other qualities such as reliability, validity, and etc.

### **2.3 TYPES OF TESTING**

There are different types of tests classified in terms of their different functions and purposes. Several researchers make the distinction between these test types clear (Baker, 1989; H.D. Brown, 2001; Cohen, in Celce-Murcia, 2001; Harmer, 2001; Heaton, 1988; Hughes, 1989; McNamara, 2000).

In the opinion of Cohen (in Celce-Murcia, 2001:516), a test can be used for twelve different purposes:

... five administrative purposes (achievement, placement, exemption, certification, promotion), four instructional purposes (diagnosis, evidence of progress, feedback to the respondent, evaluation of teaching or curriculum), and three research purposes (evaluation, experimentation, knowledge about language learning, and language use).

However, a normal task cannot include all these purposes. To Cohen, the basic distinction is often between *proficiency tests* which are usually for administrative purposes and *achievement tests* used for the assessment of instructional outcomes.

Hughes (1989) draws attention to four types of tests: *proficiency, achievement, diagnostic and placement tests*. Similarly, Harmer (2001) attaches great importance to these test types. She also names achievement tests as progress tests. H.D. Brown (2001) adds one more test type to these four basic test types which is *aptitude tests*. Heaton (1988) does not state much information about placement tests specifically; however, he suggests another name for achievement tests which is attainment tests. Finally, McNamara (2000) draws attention to the difference between proficiency and achievement tests.

Bearing all these studies in mind, five types of tests are going to be explained thoroughly in the following sections: aptitude, placement, diagnostic, proficiency and achievement (progress and attainment) tests.

### **2.3.1 APTITUDE TESTS**

Aptitude tests are given to students before any exposure to language. These tests predict students' future performances. They are designed "to measure the student's *probable* performance in a foreign language which he or she has not started to learn" (Heaton, 1988:173).

Aptitude testing requires very much effort. To some specialists, "it is neither possible nor desirable to take an overall measurement of language aptitude" (Heaton, 1988:173). Students' strengths and weaknesses are measured by means of an

artificial language. As it is really complex and as few teachers deal with aptitude testing, it is not necessary to explain it in detail here.

### **2.3.2 PLACEMENT TESTS**

As the name suggests, placement tests are used to place students into different classes or groups according to their levels. This type of tests is generally used before some private courses, or at the very beginning of the preparatory classes of the universities. For example, YADEM (Foreign Language Education Practice and Research Centre) applies a placement test to the course participants and places the participants into five different levels: elementary, pre-intermediate, intermediate, upper-intermediate, and advanced. Similarly, at COMU, preparatory classes have two different levels at the beginning of the term: elementary and pre-intermediate. Students take a placement test and are classified into different levels.

Placement test should be prepared by the institution which needs it. If there is no chance to prepare one, it can be bought. However, before buying a placement test, the institution must be sure that the test suits the programme that it aims to test (Hughes, 1989). It is really important that placement tests should be representative of the course programme. Otherwise, students' levels can be different from what they have to be. Thus, students can suffer from misplacement and this may result in the loss of a term, a year or etc. When they are in the right level, this will motivate and encourage them to move to upper levels.

### **2.3.3 DIAGNOSTIC TESTS**

In their book named "*Dictionary of Language Testing*", the term diagnostic test is defined as a test which is "used to identify test takers' strengths and weaknesses by testing what they know or do not know in a language, or what skills they have or do not have" (Davies et al., 1999).



While placement tests are designed to test students' English levels before the course, diagnostic tests aim to "expose learner difficulties, gaps in their knowledge and, skill deficiencies during a course" (Harmer, 2001:321). To Hughes (1989), it is really unfortunate not to have effective diagnostic tests. By means of these tests, learners can see their weaknesses and the teachers can take the appropriate remedial action for those weaknesses.

#### **2.3.4 PROFICIENCY TESTS**

Proficiency tests are designed to test language performances of students with different language training backgrounds (Alderson et al., 1995). Generally, these tests consist of standardized multiple-choice items on grammar, vocabulary, reading comprehension, etc. and they look to the "future situation of language use without necessarily any reference to the previous process of teaching" (McNamara, 2000:7).

To Harmer (2001:321), proficiency tests provide "a general picture of students' knowledge and ability". For example, to assess students' English knowledge and ability before entering a master programme, YADEM applies a proficiency test for the candidates. Unless the students get 40 out of 100, they cannot start their education at a master programme. As it is known by many people, TOEFL (Test of English as a Foreign Language) can be given as another example of proficiency tests. Such proficiency tests are invaluable part of the educational system in Turkey; thus, they should be carefully prepared and applied.

#### **2.3.5 ACHIEVEMENT TESTS**

Achievement tests measure what students have learned at the end of the teaching process. Unlike proficiency tests which look to the future situation of language use, achievement tests measure what students learned in the past (McNamara, 2000). These tests are directly related to *the language courses*. As this study is about students' and teachers' perceptions of the test techniques used at

*Compulsory English Language Courses*, achievement test will be the main concern of this research.

The main purpose of achievement tests is to “establish how successful individual students, groups of students, or the courses themselves have been in achieving [the course] objectives” (Hughes, 1989:10). Achievement tests, also called as progress tests, simply measures the progress of the students at a language course. Similarly, Salkind (2006:204) states some other purposes of achievement tests. All of these purposes work hand in hand and serve each other.

1. Achievement tests help define the particular areas that teachers believe are important to assess.
2. Achievement tests tell teachers and testers whether an individual has accomplished or achieved the necessary knowledge to move to the next step in study.
3. Achievement tests can allow for the grouping of individuals into certain skill areas.
4. Achievement tests may be used diagnostically in that they help identify weaknesses and strengths.
5. Achievement tests can be used to assess the success of a program....
6. Finally, achievement tests inform [about what precautions should be taken for students' better learning].

Harmer (2001:321) believes that at the end of a term achievement tests “should reflect progress, not failure. They should reinforce learning that has taken place, not go out of their way to expose weaknesses”. To make the achievement tests of the Compulsory English Language Course more effective, and to reinforce foreign language learning at COMU; a deeper look at the different achievement test types is necessary. The characteristics of and the difference between teacher-made and standardized achievement tests are explained in detail in the following sections.

### **2.3.5.1 TEACHER-MADE (NON-STANDARDIZED) ACHIEVEMENT TESTS**

As the name suggests, teacher-made tests are constructed by the teacher. The teacher himself/herself tries to establish validity, reliability and etc. of the test. Unless the test is prepared by a teacher who is knowledgeable about testing, teacher-

made test can have negative washback effect. However, any teacher can prepare his/her test effectively if he/she knows the basic considerations in testing.

In the opinion of Salkind (2006:205), “there is nothing at all wrong with the teacher-made tests – they are just very situation specific and defined to suit a particular need”. As it can be understood from his statement, teacher-made tests are very specific assessment instruments. The quizzes that the teacher applies before the midterm exam are good examples for teacher-made achievement tests. The teacher defines and determines a particular need – e.g. revising tenses – and applies a quiz to the students to satisfy this need.

Although they are not usually as practical as the standardized tests, teacher-made achievement tests are also important for better language teaching and learning; however, it takes great amount of time to prepare an effective achievement test. Most of the time teachers find it easier to prepare questions on what is easy to ask than on what is important to ask (Hughes, 1989). Thus, lack of validity becomes the main problem of such tests.

### **2.3.5.2 STANDARDIZED ACHIEVEMENT TESTS**

Standardized Achievement Tests (SATs) are the main focus of this research study. At COMU, all the faculty students have “Compulsory English Language Courses” in their first year. At the end of the course, a SAT is used to assess students’ performance. As this study will be about this SAT, it will be beneficial to explain what a SAT is.

Popham (2003:125) defines it as “any assessment device that’s administered and scored in a standard, predetermined manner”. Similarly, to Bracey (1998:17, quoted in Burke, 2005:24), standardized tests are tests which are standardized in four different areas:

1. Format. The format of all questions for all the students is the same (usually, but not always, multiple choice).
2. Questions. All the questions for all the students are the same.
3. Instructions. All the instructions for all students are the same.
4. Time allotment. The time permitted to complete the test for all the students is the same.

When these areas are taken into consideration, the midterm and the final exams of the Compulsory English Language Course are completely standardized achievement tests.

SATs gained importance because of the people's lack of confidence in the schools and the teachers. Wiggins (1989:42, quoted in Burke, 2005: 27) feels that:

standardized testing evolved and proliferated because the school transcript became untrustworthy. 'An A in *English* means only that some adult thought the students' work was excellent. Compared to what or whom? As determined by what criteria? In reference to what specific subject matter?'

All these questions made teacher-made tests untrustworthy and standardized tests started to be used more frequently.

Another reason why standardized tests are used instead of teacher-made tests is their practicality. Instead of making up your own tests, it is always easier to use a test which already available. Furthermore, "the procedures for administering, scoring, and interpreting the test have already been worked out" (Genesee & Upshur, 1996:233).

Besides their practicality, the scoring of standardized achievement tests is usually objective. Heaton (1988) believes that both subjective and objective test items should be available in a good achievement test. Moreover, he presents some examples of objective test items (multiple-choice, transformation, completion or fill in the blanks, combination, addition, rearrangement, correct/incorrect or true/false, short answer and matching) and adds that although these objective test items are usually used to assess grammar knowledge, they can equally and successfully be

used to assess vocabulary knowledge and other aspects of reading, writing, listening and speaking skills.

For fair and practical testing, standardized achievement tests are inevitable; however, “neither standardized tests alone nor teacher [-made] assessments alone can provide a true picture of student’s learning” (Burke, 2005:33). If teachers want to assess effectively, they need to use both classroom (teacher-made) tests and standardized tests. Most important of all, there must be a balance between teacher-made and standardized tests. That is to say, there must be a close relationship between the test techniques used both at teacher-made and standardized tests. What has been done at the classroom by the teacher and what is being tested at the end of the course should be in correspondence with each other.

In the following section, the research studies on the students’ and teachers’ perceptions of the test techniques used to assess language performance will be mentioned. These studies will help the researcher determine the questions of the questionnaire that will be used for this thesis.

## **2.4 TEST TECHNIQUES USED TO ASSESS LANGUAGE PERFORMANCE AT UNIVERSITY LEVEL: STUDENTS’ AND TEACHERS’ PERCEPTIONS**

Recently many studies have been carried out regarding the teachers’ and testers’ perceptions of the test techniques used to assess language performance. However, the literature on students’ perceptions of the same issue is quite limited. As the teachers’ and students’ perceptions of the test techniques will be examined in this research, several studies will help the researcher to a great extent create his own questionnaires and determine the test techniques for objective testing of students’ language performance.

Dalyan (1990) conducted a research which focused upon the assessment of teachers’ opinions on the English language testing system at the Faculty of

Engineering, Anadolu University. He aimed to establish some suggestions for more effective ways of testing at this faculty. Therefore, he asked his colleagues what they are thinking about the testing system at that faculty of Anadolu University. His study is a really good guide for this research as he suggested some techniques that can be scored objectively such as: multiple-choice, true-false, matching, rearrangement, addition, transformation, short-answer, and fill in the blank. According to the findings of his study, the most appropriate test items in the opinions of the teachers are respectively: multiple-choice, matching, true-false and fill in the blanks. He found out that most of the teachers always use multiple-choice test items to assess students' language performances. This is also the case at COMU. Only multiple-choice test items are used to assess students' success at standardized achievement tests as they are part of objective testing.

In his article named "*Vocational Education Teachers' Perceptions of Their Use of Assessment Methods*", Gordon (1998) provides a questionnaire consisting of "semantic differential scales" (Dörnei, 2003:39). He asks the teachers to rate whether the assessment methods are valuable or worthless, successful or unsuccessful, efficient or inefficient, etc. by means of a scale. In the light of this scale, it is aimed to prepare a similar one and find the students' and teachers' perceptions of the present testing system at COMU.

In his research, Şahinel (1997) aims to determine English lecturers' opinions on the English language testing situations at the preparatory classes of Ankara University. He obtained the data by means of a questionnaire and he found that the test techniques used at the exams had not been arranged in the order of difficulty and the objectives of English language curriculum had not been taken into consideration by testers while constructing their tests. Taking all these findings into consideration, I planned to identify whether the test content is reflecting the course objectives or not at *Compulsory English Language Courses*, at COMU.

In addition to these studies; Aksan (2001) at Niğde University, Ösken (1999) at Hacettepe University and Serpil (2000) at Anadolu University administered

questionnaires to the instructors and found out their perceptions of the content validity of the English language tests. The questions they asked to the instructors in their questionnaires shed light on constructing the questionnaires of my study.

In her thesis, Kuntasal (2001) aimed to find out teachers' and testers' perceptions of achievement tests prepared by testers in the department of Basic English at the Middle East Technical University. Her thesis helped me determine the questions of my questionnaire, especially the ones about what the instructors think about the practices of the testing office at COMU. Furthermore, she suggests other researchers to find out what the students think about the achievement tests for further research. Taking this suggestion into consideration, it is also aimed to find out the students' perceptions of the test techniques used by testers of COMU.

Moreover, in Vergili's (1984) master thesis entitled "*Teachers' Attitudes toward Testing at METU Gaziantep Preparatory School*", there were one tester and fifteen teachers as participants. They were all given a questionnaire and their answers were compared. In the conclusion part, Vergili claims that there is a lack of coordination between testers and teachers. This claim raises important questions for my study: "Is there a lack of coordination between the testers of the testing office and the instructors at COMU? Are the instructors happy with the current practices of the testing office?"

Urduan and Paris (1994) also investigated "*Teachers' Perceptions of Standardized Achievement Tests*". To them, it is really important to understand to what teachers think about the standardized achievement tests as these thoughts may determine the ways how they prepare and administer the test practices. Also Cohen (in Celce-Muricia, 2001:515) claims that students and teachers are afraid of the word testing. Students are afraid of tests because they think that they will not perform well. As for the teachers, "they do not construct tests and are not altogether satisfied with the results when they do. They are also suspicious of the standardized ... tests because they are not always sure what these tests are actually trying to measure" (Cohen, in Celce-Muricia, 2001:515). His claim raises other important questions to

be answered: is this the same case at COMU, what do the instructors think about the SAT prepared by testers and are they satisfied with the works of the testing office?

In Rudman's (1989) article, teachers' and students' attitudes towards the use of tests are explained. To him, teachers who are more experienced are more supportive for the use of tests than are those who are less experienced and less knowledgeable. As for the students, they feel that "frequent testing helps them retain more content, reduces test anxiety, and aids their own monitoring of their process" (Rudman, 1989:4). These findings raise other important questions: "What do the students and the instructors think about the exams prepared by the testers of the testing office? Do they prefer the exams to be prepared by the test constructors or by the instructors of the course?"

As this study is not only about students' and instructors' perceptions of the practices of the testing office but also about the test techniques used to assess students' language performances at SATs, the literature on the objective test techniques will be given. Later, the advantages and disadvantages of these techniques will be explained in detail below. Furthermore, some suggestions were made for the scoring process of the subjective tests to be more efficient following the objective test techniques for SATs.

#### **2.4.1 OBJECTIVE TESTS FOR STANDARDIZED ACHIEVEMENT TESTS**

The terms *objective* and *subjective* are used for the scoring of tests. A test can be called as objective or subjective by determining the way how the teacher scores the students' performances. Subjectiveness can be used for everything. Not only the scoring but also the construction of tests can be subjective as the testers ask whatever they want to ask. However, "only the *scoring* of a test can be described as objective" (Heaton, 1988:25). To be called as an objective test, a test must "have only one correct answer (or, at least, a limited number of correct answers)" (Heaton, 1988:25).



At COMU, Compulsory English Language Courses, only multiple-choice test items are used to assess students' language performance. To Köksal and Beceren (2003), multiple-choice tests may not be completely enough to assess students' ability to produce. Multiple-choice question form used at COMU only tests the ability to recognize correct grammar and usage. However, they never test the ability to produce. This study aims to review the literature on the test techniques that can both be scored objectively and test students' ability to produce to some extent.

Köksal (2004) believes that while choosing their tasks in tests, teachers should take the instructional activities used in the classroom into consideration. At Compulsory English Language Courses, instructors use many kinds of activities consisting of different test techniques while following their coursebooks. Nevertheless, as it is a SAT, they can only test students' performance by multiple-choice test technique in the final exam. What techniques the instructors use in the class and what techniques the testers should use in the SAT of Compulsory English Language Course should be in consistency. This consistency will be questioned in this thesis.

While conducting a research on EFL teachers' effectiveness in testing and evaluating students' performance, Bostan (2005) provided many kinds of tests techniques to assess different language skills. She explained the advantages and disadvantages of all these test techniques. Among them; true-false, multiple-choice, short answer, the cloze test, C-test, cloze elide test, gap filling and ordering tasks can be regarded as tests which are scored objectively. Thus, they can be used in standardized tests.

Kitao and Kitao (1996) also explain the advantages and disadvantages of the test techniques used to test grammar. In addition to testing the ability of recognition by means of multiple-choice questions, they provide other question forms such as error correction, sentence/word order, transformation items, word changing items, and sentence combining items to test the ability to *produce* correct grammar and usage.

In their article titled *Using Objective Tests to Evaluate*, Parsons and Fenwick (1999) argues that the testers should use different types of questions for objective tests such as fill-in-the-blank, short-answer, multiple-choice, true-false and matching. They believe that objective tests can take a lot of time while being prepared, but they are relatively easy to grade. Similarly, Frost (2005) provides the advantages and disadvantages of different test techniques: multiple-choice, transformation, gap filling, matching, cloze and true-false. In his thesis, Dalyan (1990) also lists some objective test techniques as multiple-choice, true-false, matching, rearrangement, addition, transformation, short-answer and fill-in-the-blank.

Moreover, Madsen (1983) gives many examples of the techniques used in language testing. Like Madsen, many authors (Heaton, 1988; Genesee & Upshur, 1996; Hughes, 1989; Alderson, 2000; Weir, 1990) provide examples of similar test techniques used to assess language performance. In the light of this literature, it is aimed to explain the advantages and disadvantages of some techniques that can be used at standardized achievement tests for *objective testing*. Before the methodology part, all of the *objective test techniques* [(1) Multiple-choice, (2) Short-answer, (3) True-false, (4) Matching, (5) Completion, (6) Cloze test, (7) C-test, (8) Cloze elide tests, (9) Ordering tasks (Rearrangement), (10) Error correction, (11) Transformation, (12) Combination and Addition, (13) Word changing] that were found during the literature review will be given in detail in the following sections.

#### **2.4.1.1 MULTIPLE-CHOICE QUESTIONS**

Multiple-choice tests are probably the most common way of testing grammar knowledge. However, some other language skills and areas can also be tested by means of this test technique. The suggestions for the construction of multiple-choice test items are applicable to testing listening, vocabulary, grammar, reading and pronunciation. This test technique can be very useful for testing some skills and areas, but will not be useful for testing all language skills. Madsen's (1983) statement

makes this fact clearer. To him, “while multiple-choice tests can be used successfully in testing grammar, they do not seem to work as well in testing conversational ability” (p.38).

Multiple-choice questions take many forms. Hughes (1989:59) points out that “there is a *stem* and a number of *options*, one of which is *correct*, the others being *distracters*”. The teacher can give the question either through an incomplete sentence or through a full question. Following example demonstrates the things that make a multiple-choice question.

**Stem** → Do not phone your father until he \_\_\_\_\_ you.

<b>Options / Responses /← Alternatives</b>	a) phoned	<b>→ Distracters</b>
	b) will phone	
	c) is phoning	
	d) phones	<b>→ Answer / Correct Option / Key</b>

(Adapted from Heaton, 1988)

Besides grammar, the teacher can test students’ knowledge of pronunciation by means of multiple-choice test items. In his article on *Testing Pronunciation*, Bobda (1993) gives an example as follows:

In each of the following sets of words, three words have the same sound and one does not. Circle the one that does not have the same sound with the others.

1. a) dull                      b) bull                      c) wool                      d) pull
2. a) poor                      b) pour                      c) sure                      d) tour

Multiple-choice item seems to be constructed very easily. However, it is one of the most difficult and time-consuming type of item to construct (Heaton, 1988; Hughes, 1989; Madsen, 1983). Thus, teachers should take these basic principles while constructing their multiple-choice test items (Hughes, 1989).

1. A multiple-choice question should have only one correct answer.
2. The teacher should test one thing at a time.
3. Each option should be grammatically correct when placed in the stem.

Someone who flies a plane is a \_\_\_\_\_.

- a) architect                      b) pilot                      c) teacher                      d) accountant

A student who answers this question will soon understand that the options “a” and “d” are not the correct answer as they form grammatically incorrect sentence when placed in the stem. The teacher can write “a / an” in the stem to get rid of this ambiguity.

4. Multiple-choice questions should be at an appropriate level to the proficiency level of the students.
5. They should be brief and as clear as possible.
6. Finally, these items should be ordered in terms of their difficulty level. Questions should be given from the easiest one to the most difficult one. At least, having two or three easy questions at the beginning of a test will be a good chance for the students to ‘lead-in’.

This technique has its advantages and disadvantages. Awareness of this technique’s strengths and weaknesses is an important factor for the teachers’ success in evaluating their students’ performance.

### **Advantages**

1. Marking multiple-choice test items is not only reliable but also simple (Hughes, 1989; Madsen, 1983; Salkind, 2006; Weir, 1990).
2. These items can be pre-tested easily. As a result of pre-testing, ambiguities in the test design can be clarified or removed before the test is applied to the students (Weir, 1990).
3. They can be used to “measure learning outcomes at almost any level” (Salkind, 2006:140).
4. To Alderson (2000:211), “they allow testers to control the range of possible answers to comprehension questions”. But for multiple-choice test items, there would be many possible answers given by the students during their reading exams and the scoring of these answers would be almost impossible.

### **Disadvantages**

1. It takes a long time to write effective multiple-choice test items (Alderson, 2000; Heaton, 1988; Hughes, 1989; Madsen, 1983; Salkind, 2006).

2. It is easy for students to cheat. Cheating is facilitated as there are only four or five options. Students can communicate with their friends non-verbally and give the correct answer using their body language (Hughes, 1989; Madsen, 1983).

Hughes (1989) provides other disadvantages of multiple-choice test technique:

3. This technique tests only receptive knowledge, not productive skills.
4. Guessing may have a considerable but unknowable effect on test scores.
5. This technique restricts what can be tested as it is really difficult to find enough distracters for the correct structure to be tested.
6. Backwash may be harmful. In some cases, students are trained in guessing rather than in learning the language. Thus, multiple-choice tests have a harmful effect on learning and teaching.

#### 2.4.1.2 SHORT-ANSWER

In this technique, students are asked to write down specific answers in spaces provided on their exam papers. This technique is very useful for the testing of reading and listening. Short-answer questions “give students a structure for answering, but allow more freedom [than multiple-choice questions] in these answers” (Parsons & Fenwick, 1999:5).

Gronlund (1971:150) highlights that “short-answer test items which can be answered by a word, phrase, number or symbol are constructed by direct questions”. Following example can be a good one for a short-answer test technique:

Answer the question below. Use full sentence in your answer.

“When was your mother born?” \_\_\_\_\_.

There may be many different years in students’ answers, but the structure of “*She (My mother) was born in ...*” will be the only accepted answer. Thus, it will be nice to use such questions both for objectivity of the scoring and for the students’ productive answers.

The answers to such direct questions cannot objectively be scored unless the tester takes these principles into consideration while constructing short-answer test items (Parsons & Fenwick, 1999):

- The teacher should avoid using an open-ended question if he/she wants to determine students' knowledge of a particular area. He/She should ask questions that need to have one or two predetermined answers.
- The teacher should be very clear and specific in his/her wording.

Ambiguous questions may cause long answers. Thus, the grading becomes more difficult and more subjective. An effective teacher of English should make use of this technique's advantages and be very careful about its disadvantages. Weir (1990) emphasizes the advantages and disadvantages of short-answer test items as follows:

### **Advantages**

1. Answers are not provided for the students as in multiple-choice. When the student gives the right answer, the teacher can be sure of his/her knowledge as there is no chance to guess as in multiple-choice.
2. If the teacher formulates the questions carefully, students' responses can be brief. Thus, the teacher can ask a large number of questions which helps him/her to cover most part of the content to be asked.
3. If the number of accepted answers to the questions is limited, it will be fairly scored by any markers.
4. This type of questions can provide more reliable data about students' reading ability.

### **Disadvantages**

1. This technique requires the students to write. Thus, this interferes with the measurement of intended construct. Is "writing" or "some other construct of a language" being tested when using this technique?
2. If the teacher does not limit the range of possible acceptable answers, there is a "possibility that the variability of answers may lead to marker unreliability" (Weir, 1990:45).

### 2.4.1.3 TRUE-FALSE QUESTIONS

As another measuring instrument, true-false test item is extensively used in the EFL classroom. In the opinion of Smith and Adams (1972:147), “this group consists of any question in which the student is confronted with two possible answers”. This technique can be used by presenting students with a graph, chart or a text. Students can be asked to rate the statements true or false based on their understanding the graph, chart or text. The example below clarifies the description of true-false test items.

Look at the information in the chart and decide if the sentences are true (T) or false (F).

#### WORLD CUP UPDATE

Morocco – Tunisia	___	USA – Germany	in progress
S. Korea – Morocco	___	Japan – Australia	in progress
Croatia – England	0 – 0	Australia – USA	3 – 0
Tunisia – Croatia	3 – 1	Germany – Japan	0 – 0
Morocco – England	2 – 1	Australia – Germany	1 – 1

- |  |   |   |
|--|---|---|
| 1. England have already played Morocco.      | T | F |
| 2. Tunisia haven't played Croatia.           | T | F |
| 3. Japan still haven't played a match.       | T | F |
| 4. The United States have already lost once. | T | F |
| 5. South Korea haven't played yet.           | T | F |

(Taken from Thornbury, 1999: 144)

In order to use this technique effectively, in the opinion of Parsons and Fenwick (1999), teachers should avoid using generalizations as they are not always 100 per cent true. They should use statements which are completely true or false. Furthermore, they should not use negative statements as these statements can be confusing for the students. Also, the statements should not be very long and should not contain two ideas.

Heaton's (1988) opinions about the advantages and disadvantages of true-false tests are well-founded. He believes that:

### **Advantages**

1. The scoring of these tests is quick and the scores can be very reliable if the items are well constructed.
2. They are useful for class progress tests as, unlike multiple-choice items, they can be constructed easily and quickly.
3. Their quick construction allows the teacher more time for other tasks.

### **Disadvantages**

1. This technique encourages guessing as the students have 50% chance of giving the right answer for each question. Heaton (1988) suggests two different solutions to this “guessing” problem. Firstly, at the beginning of the test, the teacher can add instructions similar to the one below.

Each correct answer will be awarded two marks. However, for each wrong answer, one mark will be deducted from your score. It is better, therefore, not to guess blindly and to leave a blank if you do not know the correct answer (p.114).

Secondly, the teacher can add “a third question in addition to the true/false options: e.g. true, false, not stated” (Heaton, 1988:114).

2. Unless there are a lot of true-false questions, the test may not be able to discriminate students’ success.

#### **2.4.1.4 MATCHING**

Matching test items are just like multiple-choice test items in many ways. Salkind (2006:157) treats them as the same because “matching items are multiple-choice questions that share the same alternatives”. The example below is a good example of this technique. However, it can be better if it is accompanied by a text.

<u>Options</u>	<u>Premises</u>
___ 1. bang	A- a woman member of religious community
___ 2. gallon	B- to cry suddenly taking sharp breaths
___ 3. nun	C- a unit for measuring liquid
___ 4. sob	D- to receive property from someone when they die
___ 5. inherit	E- to hit something in a way that makes a noise



Here are some guidelines for constructing effective matching items (Salkind, 2006:160-162):

1. There are two columns, one containing premises and the second containing options or responses...
2. As with any tests, provide complete directions. ... Let the test taker what is allowed and supposed to be done...
3. All premises and all responses should be reasonable...
4. Responses should not be listed in the same order as the corresponding premise...
5. A premise should contain more words than an option...
6. To help all test takers respond efficiently, place the premises in some logical order, such as alphabetical...
7. Make sure that all premises and responses appear on the same page...
8. Make sure that each premise has only one correct response.

After stating the guidelines for how to write matching items, it is time to look at their advantages and disadvantages. Salkind (2006:163) summarizes the advantages and disadvantages of these items as:

#### **Advantages**

1. They are straightforward and clear in their presentation.
2. They are easy to administer and responses to these tests are short and easy to read.
3. They allow for comparison of ideas and facts.
4. The value of guessing is decreased.

#### **Disadvantages**

1. The level of knowledge tested is limited and the test emphasizes memory.
2. The test can be difficult to machine score.

#### **2.4.1.5 COMPLETION**

This technique is also called as ‘fill-in-the-blank’ or ‘gap filling’. Although this kind of test technique seems to be scored subjectively, “they are often regarded as belonging more to the objective category of test items” (Heaton, 1988: 124).

Completion test items can be used to test different areas such as vocabulary, grammar and etc.

Like short-answer test items, completion items are easy to construct. However, while constructing these items, testers should be careful. They should avoid ambiguous sentences that can be completed in several different ways. The teacher can provide set of words below the questions for students to choose and ask them to fill in the blanks with the words they have chosen. This changes completion items into matching ones. However, it can be a good solution for this ambiguity (Parsons & Fenwick, 1999). Following test by Thornbury (1999:143) exemplifies this test technique:

Complete this text with *yet*, *still*, or *already*.

Preparations are underway for the Pan-World Games in Lomoka next year. Many new hotels have (1) \_\_\_\_\_ been built and tourists are (2) \_\_\_\_\_ making reservations. But the main stadium hasn't been started (3) \_\_\_\_\_. They are (4) \_\_\_\_\_ deciding where to put it. The Athletes' Village is (5) \_\_\_\_\_ being built, and the swimming complex isn't completed (6) \_\_\_\_\_.

### **Advantages**

1. This type of question is easy to write and mark.

In the opinion of Madsen (1983):

2. Completion test technique seems to measure productive skills as they allow some extent of flexibility in students' answers.

3. Students' focus on the correct grammar form.

### **Disadvantages**

1. Although it is easy to construct such items, it is difficult and time-consuming to correct students' mistakes. Irrelevant errors can be a problem for the markers (Madsen, 1983).

2. As Hughes (1989:150) suggests, there can be more than one possible correct answer. He advises the teachers to provide students with "the first letter of the word (possibly more) and even an indication of the number of letters [when assessing vocabulary]".

### 2.4.1.6 CLOZE TEST

Cloze tests are effective ways of testing grammar, vocabulary and reading. When prepared carefully, they can be scored objectively; thus, can be used at standardized achievement tests. “A true cloze test is a text in which every  $n$ th word has been deleted” (Thornbury, 1999:145). It can be every seventh, or every ninth word, etc. This number ( $n$ ) is usually between five and twelve. Below, there is a good example of a cloze test in which every seventh word was deleted. In such tests, every “student relies on the context in order to supply the missing words” (Madsen, 1983:47).

Preparations are underway for the Pan-World Games in Lomoka next year. Many new hotels have (1) \_\_\_\_\_ been built and tourists are already (2) \_\_\_\_\_ reservations. But the main stadium hasn't (3) \_\_\_\_\_ started yet. They are still deciding (4) \_\_\_\_\_ to put it. The Athletes' Village (5) \_\_\_\_\_ still being built, and the swimming (6) \_\_\_\_\_ isn't completed yet....

**(Taken from Thornbury, 1999: 145)**

Madsen (1983:48) gives some principles that have to be taken into consideration during the construction of cloze tests: “(1) select an appropriate passage (e.g. from the reading material in your ESL class); (2) decide on the ratio of words to take out; (3) write the instructions and prepare an example”. He also claims that these tests are advantageous, but have some limitations as well:

#### **Advantages**

1. These tests are easy to prepare and quite easy to score.
2. They are good ways of testing integrative English skills.
3. They are effective at testing overall ability in English.

#### **Disadvantages**

1. This test technique is a good one to test long-term gains as it tests overall ability in English. Thus, it is not a good measure of short-term gains.
2. Teachers, especially who are non-native English speakers, can find it difficult to choose acceptable equivalent words.

### 2.4.1.7 C-TEST

This test technique is similar to the cloze tests. However, instead of whole words as in cloze tests, second half of every second word is deleted in this technique. Raatz (1985:17; quoted in Cohen, in Celce-Murcia, 2001:521) provides an example of C-test as follows:

Pollution is one of the big problems in the world today. Towns a\_\_ cities a\_\_ growing, indu\_\_ is gro\_\_ and t\_\_ population o\_\_ the wo\_\_ is gro\_\_. Almost every\_\_ causes poll\_\_ in so\_\_ way o\_\_ another. T\_\_ air i\_\_ filled wi\_\_ fumes fr\_\_ factories a\_\_ vehicles, a\_\_ there i\_\_ noise fr\_\_ airplanes a\_\_ machines. Riv\_\_, lakes, a\_\_ seas a\_\_ polluted b\_\_ factories and by sewage from our homes.

#### Advantages

1. C-tests are more reliable than cloze tests (Alderson, 2000; Weir, 1990).
2. They can be objectively scored.
3. Hughes (1989:71) claims that when compared with cloze tests, “a C-test of 100 items takes little space and not nearly so much time to complete”.
4. C-tests function well as a measure of overall ability in a foreign language (Hughes, 1989).

#### Disadvantages

1. To Hughes (1989), these tests are like a puzzle and it is more difficult to read than a cloze test.
2. Weir (1990:49) highlights that “this technique suffers from the fact that it is irritating for students to have to process heavily mutilated texts and the face validity of the procedure is low”.

### 2.4.1.8 CLOZE ELIDE TESTS

This technique is another alternative to the cloze technique. It was invented by Davies in 1960s. In its earlier times, it was known as “Intrusive Word Technique” (Davies, 1975; quoted in Alderson, 2000:225). In this test technique, the tester inserts words instead of deleting them. Students have to show where these insertions have

been made. Alderson (2000) notes that students are supposed to delete the words that do not belong to the original text. He also advises the teachers to award students with a point for every word correctly deleted and deduce for the words that are wrongly deleted. Students are supposed to underline every inserted word as it was done in the following example:

Aşure is with the name of a pudding which is been cooked with extremely diverse and seemingly disharmonious ingredients. It is a pudding through that has become a part of many the Muslim tradition. It is believed to have originated from when Noah's Ark came to rest on a mountain of the Turkey after the great floods...

(Original text was taken from Tataroğlu, 1995:44)

### **Advantages**

1. To Weir (1990), cloze elide test has approximately the same advantages as a cloze test. In this technique, students do not have the problem of understanding the question.
2. Like cloze tests, they can be objectively scored when they are prepared carefully.
3. They can evaluate students' performance at every level of reading development (Madsen, 1983).

### **Disadvantages**

1. Weir (1990:50) believes that scoring of these test items may be really problematic as students can "delete items which are correct, but redundant".
2. Madsen (1983:97) believes that this technique "encourages word-by-word reading". He also says that "passage comprehension is more time consuming to take than any other kinds of tests".

#### **2.4.1.9 ORDERING TASKS (REARRANGEMENT)**

In an ordering task, students are given "a scrambled set of words, sentences, paragraphs or texts... and have to put them into their correct order" (Alderson, 2000:219). This technique is generally used to test simple or complex grammar, cohesion, and reading comprehension. An example of an ordering task can be given as follows:

The following sentences and phrases come from a paragraph in an adventure story. Put them in the correct order. Write the letter of each in the space on the right.

Sentence D comes first in the correct order, so D has been written beside the number 1.

- |   |             |
|---|-------------|
| A it was called "The Last Waltz"                          | 1. <u>D</u> |
| B the street was in total darkness                        | 2. ____     |
| C because it was one he and Richard had learnt at school. | 3. ____     |
| D Peter looked outside                                    | 4. ____     |
| E he recognized the tune                                  | 5. ____     |
| F and it seemed deserted                                  | 6. ____     |
| G he thought he heard someone whistling                   | 7. ____     |

(Taken from Alderson et al., 1995:53)

As Alderson et al. (1995) argue, there are at least two ways of ordering this paragraph. Thus, the teachers should be careful about improving their items before directly applying them to test their students' performance. They suggest that adding "but" to the beginning of phrase G makes only one acceptable answer key: 1:D, 2:B, 3:F, 4:G, 5:E, 6:C, 7:A. Such improvements in the test items are inevitable for effective testing.

Heaton (1988:41) names ordering tasks as "rearrangement items". He believes that these items can take several forms. The tester can present the question either in a multiple-choice format or as a word-order item. Examples of these different forms can be given as in the following:

**Example of an ordering task in multiple-choice format:**

'Won't I need a coat?'

'Well, you know how \_\_\_\_\_.'

- |                     |                     |                     |
|---------------------|---------------------|---------------------|
| A) warm is it today | B) today it is warm | C) is it warm today |
| D) warm it is today | E) today is it warm |                     |

**Example of an ordering task as a word-order item:**

Complete each sentence by putting the words below it in the right order. Put in the boxes only the letters of the words.

'Won't I need a coat?'

'Well, you know how \_\_\_\_\_.'

- |       |          |         |       |  |
|-------|----------|---------|-------|--|
| A. it | B. today | C. warm | D. is |  |
|-------|----------|---------|-------|--|

--	--	--	--

(Taken from Heaton, 1988:41)

### **Advantages**

1. This technique enables the testers to test the students' ability to detect cohesion, overall text organization, or complex grammar (Alderson, 2000).
2. When prepared in multiple-choice format, they can be scored objectively.

### **Disadvantages**

1. Alderson et al. (1995:52) claim that "all ordering tasks are difficult to construct because it is not easy to provide words or phrases which only make sense in one order".

#### **2.4.1.10 ERROR CORRECTION**

Error correction items are useful for testing grammar. These tests usually consist of sentences or passages in which there are some errors that the students have to identify and correct. For example, students can be given a sentence similar to the one below:

Only one word in the following sentence is an error. Write the correction of that word in the space given beside the sentence.

1. Ali and Ayşe is going to watch the match tomorrow.            1. \_\_\_\_\_

Kitao and Kitao (1996) suggest that teachers should also provide students with some sentences that have no errors. In that case, students are supposed to indicate that there is no error. In their opinion, errors from students' own written materials are effective sources of ideas for this test technique. Learning will be facilitated when students learn from their own errors.

### **Advantages**

1. When students' own written materials are used, students will learn from their own errors.
2. Errors can be given in multiple-choice format which makes the scoring of this technique quite objective.

### **Disadvantages**

1. Alderson et al. (1995:54) assert that “the main difficulty with this kind of item is to make sure that there is only one mistake... [per sentence]”.

#### **2.4.1.11 TRANSFORMATION**

In transformation technique, the tester gives a sentence and the first few words of another sentence. Then, he/she asks students to change the original sentence without changing the meaning as it is seen in the example below:

Rewrite each of the following sentences in another way, beginning each new sentence with the words given. Make any changes that are necessary but do not change the general meaning of the sentence.

1. I haven't written you for a long time. It's a long time \_\_\_\_\_.
2. In sunny weather I often go for a walk. When the weather \_\_\_\_\_.

(Taken from Heaton, 1988:46)

Heaton (1988) believes that this technique has some advantages and disadvantages. To him:

### **Advantages**

1. It is really useful for testing ability to produce structures in the target language.
2. In multiple-choice format, it can be scored objectively.
3. This technique is effective for testing grammar (Frost, 2005).

### **Disadvantages**

1. It is often difficult to restrict the number of possible answers. Unless these items are prepared carefully, scoring of them can be really difficult for the teachers.

#### **2.4.1.12 COMBINATION AND ADDITION**

As combination and addition items involve students' mechanical responses, they are objectively scored (Heaton, 1988). Examples of these items are given below (Adapted from Heaton, 1988).



Combine the sentences using the words in brackets.

1. I met a woman. The woman had kissed me before. (Who)
2. You finish your homework. Then, check it carefully. (After)

Insert the words in capitals in the most appropriate place in each sentence.

1. YET                      Have you finished your homework?
2. ALWAYS              She helps her mother.

### **2.4.1.13 WORD CHANGING**

This type of test technique is similar to the addition. In this technique, the students are given an incomplete sentence and a word which they need to complete the sentence by changing the form of this word and by inserting it into the sentence in its correct form. Following examples makes this technique clearer.

Change the forms of the words in brackets and complete the sentences with the words in their correct form.

1. I have never \_\_\_\_\_ Chinese food. (eat)
2. The flowers should be \_\_\_\_\_ everyday. (water)
3. He is sleeping \_\_\_\_\_ . (moment)

In the opinion of Kitao and Kitao (1996), this test technique tests students' knowledge of various word forms and how these word forms are used in sentences.

### **2.4.2 SUBJECTIVE TESTS AS OBJECTIVE ONES**

As Salkind (2006) claims, subjective tests are not easy to score. He (2006:126) believes that all the teachers "will have some serious problems remaining neutral, staying on task, and being consistent" while scoring subjective tests. Thus, these tests are not much suitable for SATs. However, in order to make scoring process of the subjective test techniques more efficient, Salkind (2006: 126-128) suggests the following facts:

1. Scorers should provide plenty of time to score such tests.
2. Scores should grade exams in batches (perhaps one question at a time).

3. Scorers should use a model correct answer to have a basis for comparison.
4. Scorers should score each question across all test takers. For example, score the answer to Question 1 across all test takers, and then go back and score Question 2 across all test takers.
5. If possible, grade the responses without knowing the test taker's identity.

Taking all these suggestions into consideration, some of the subjective test techniques can also be used in SATs as objective tests. However, the test constructors should be really careful while constructing and the instructors while scoring such tests.

## **2.5 SUMMARY**

This chapter mainly aimed to explain the basic issues about testing and to review the studies on students' and teachers' perceptions of the test techniques used to assess language performance.

After explaining the terms - testing, evaluation, measurement, language teaching and language testing; characteristics of good tests are presented in detail. Then, types of testing (aptitude, placement, diagnostic, proficiency, and achievement tests) are mentioned. Later on, the literature on the students' and the instructors' perceptions of the test techniques used at standardized achievement tests is provided. Finally in this chapter, the advantages and disadvantages of the test techniques that can be used at standardized achievement tests are stated.

## **CHAPTER THREE**

### **METHODOLOGY**

#### **3.0 INTRODUCTION**

This chapter is about the methodology used in the study to find out the students' and the instructors' perceptions of the exams used to assess language performance at university level. In the first part, how the questionnaire was constructed and what parts it consists of are explained in detail. Second part of the chapter is mainly about the administrating the questionnaires. In order to avoid any difficulties in the main study, how piloting was done is presented in this part. Finally, main study with its setting, participants, materials, procedure for data collection and analysis is presented in the last part of this chapter.

#### **3.1 RESEARCH METHOD USED IN THE STUDY**

*Quantitative research methodology* was used in this descriptive study. Being a sub-category of a *survey method*, two questionnaires were prepared by the researcher. Questionnaires used in the study are explained in the following parts. The construction process and the description of the questionnaires are presented in detail.

##### **3.1.1 CONSTRUCTION OF THE QUESTIONNAIRES**

After reviewing the literature, the items of the questionnaires were constructed. Firstly, both questionnaires consisted of several items to be answered by the students and the instructors. Before applying the questionnaire for the piloting, the ideas of three experts on 'English Language Teaching' and one expert on 'Measurement and Evaluation' were asked. In the light of their ideas, some of the items were changed and some others were omitted. Also, some corrections were

made in the questionnaires by the experts. Then, different parts of each questionnaire were organized in order to find answers to each research question of the study.

The researcher constructed many parts of the questionnaire in the light of the literature review. However, the third part of each questionnaire was constructed by taking 73 students' and 9 instructors' ideas into consideration. Whether the testing office or the instructor himself/herself should organize the exams of the Compulsory English Language Course was asked by the researcher in an open-ended question form. The students and the instructors were asked to specify the reasons why they prefer either testing office or the instructor to organize the exams in these open-ended questions (see Appendix A and Appendix B for the open-ended questions).

When the reasons that were specified by 73 students are examined, it can be clearly seen that the students expressed their reasons for preferring either the testing office or the instructor to organize the exams under these five headings:

1. Differences in students' levels
2. Test techniques used in the exams
3. Fairness
4. Reliability
5. Instructor's attitude towards his/her students

These headings helped the researcher write the items of the third part of the questionnaire. Five items regarding the common headings that were specified by the students were constructed for each of choice – (1) testing office or (2) the instructor. Thus, one can easily see the parallelism among the items in the third part of the students' questionnaire (see Appendix D).

As for the instructors' questionnaire, the same procedure was followed. Instructors expressed their reasons for their choice under these common headings:

1. Language teaching materials
2. Curriculum
3. Test techniques used in the exams
4. Reliability
5. Differences in students' levels

After determining the five common headings listed above, the researcher constructed five items accordingly for each choice - (1) testing office or (2) the instructor. Similarly, as the items were constructed taking these five common headings into consideration, there is a parallelism among the items in the third part of the instructors' questionnaire as in the students' questionnaire (see Appendix E for the instructors' questionnaire).

Review of the literature and the ideas of both the students and the instructors helped the researcher construct the questionnaires. Finally, the need for the explanations of some terms urged the researcher to construct a glossary part for both questionnaires.

Having constructed all parts of the questionnaires, the researcher conducted the pilot study with 33 instructors and 121 students. After consulting the ideas of the instructors in the pilot study, the researcher omitted some parts of the questionnaires and added some extra items. As it was seen that students do not read the glossary part when it was given separately, the researcher inserted the explanations of the objective test techniques into the table on which the students tick their answers (See Appendix D; Students' Questionnaire, Part 5).

Before implementing the pilot and the main study, the final form of the questionnaires were constructed. The parts of these questionnaires, what they include and what they aim to measure are described in the following part in detail.

### **3.1.2 DESCRIPTION OF THE QUESTIONNAIRES**

In order to answer the research questions, two questionnaires were constructed by the researcher. Though they have many similarities they differ in terms of the samples that they were constructed for. Both questionnaires – students' and instructors' – consist of five common parts. However, 'Glossary' part was given separately in instructors' questionnaire as the sixth part while the explanations of the

objective test techniques were inserted into the table in the fifth part of the students' questionnaire (See Appendix D and E for the questionnaires).

In each questionnaire, there is an introduction section in which the purpose of the research is explained. Before the first part of the questionnaire, the researcher promises the participants that the information gathered from the questionnaires will be kept confidential so that the samples will not feel obliged to write their names on the questionnaires.

Following the introduction part, personal information about the samples was asked in the first part. Each instructor's gender, teaching experience, educational background, graduation department and knowledge on 'Testing and Evaluation' are asked in this part. As this study aims to find out the perceptions of the instructors who have only taught at faculties or colleges, there is one more question in this part asking whether the instructors have taught at a faculty or a college.

As for the students' questionnaire, students' gender and their faculties or colleges were asked in order to make stratified sampling of the whole population of the second year students. Also, students' performance at 'Compulsory English Language Course' exams is aimed to be identified in the first part of the questionnaire.

In the second part of the questionnaires, it is aimed to find out which language skill or area that the students expect to be developed most and which one the instructors try to develop most in their lessons. Are there any significant differences between the language skills and areas that the students expect to be efficient at and the ones that the instructors use in their classroom activities? The answer to this question is supposed to be found in the second part of each questionnaire. The students and the instructors are to put a cross (x) into the most appropriate box for themselves. Four language skills and three language areas were asked using Likert-type scales (5=Always; 1=Never).

Third parts of the questionnaires aim to identify students' and instructors' ideas on whether the exam should be prepared by the testing office or by the instructor of the course himself/herself. Students' and instructors' most common reasons for preferring either testing office or the instructor of the course for the preparation and the organization process of the exam are aimed to be identified in this part. Students and instructors express their ideas by putting a cross (x) for the most appropriate reason ranging from '5=Strongly Agree' to '1=Strongly Disagree'.

There is a parallelism among the reasons specified in the third parts of the questionnaires. Five reasons are presented for each choice (testing office or the instructor) under these five common headings given below.

*Table 1: Third Part of the Students' Questionnaire*

<b>Who should organize the exams of the Compulsory English Language Course?</b>	
<b>Testing Office</b>	<b>The Instructor of the Course</b>
<i>Differences in Students' Levels</i>	
1. Centrally administered achievement tests determine the differences in students' levels better.	1. The instructors prepare their questions by taking their students' different language levels into consideration.
<i>Test Techniques Used in the Exams</i>	
2. Multiple-choice test technique is used in the exams.	2. Some other test techniques can also be used in the exams.
<i>Fairness</i>	
3. Centrally administered achievement tests measure students' performance fairly.	3. As there are differences in students' levels, it is fairer that the instructors prepare their own exams.
<i>Reliability</i>	
4. Centrally administered achievement tests are more reliable.	4. The exams prepared by the instructor of the course are more reliable.
<i>Instructors' Attitude towards his/her Students</i>	
5. Centrally administered achievement tests hinder the instructors' possible negative attitudes towards their learners.	5. Instructors' possible positive attitudes towards their learners affect the evaluation process positively.

Similar parallelism can also be seen in the third part of the instructors' questionnaire as shown in the following table.

*Table 2: Third Part of the Instructors' Questionnaire*

<b>Who should organize the exams of the Compulsory English Language Course?</b>	
<b>Testing Office</b>	<b>The Instructor of the Course</b>
<i>Language Teaching Materials</i>	
1. All the instructors should follow the same language teaching materials.	1. I want to use extra language teaching materials and prepare my own exam accordingly.
<i>Curriculum</i>	
2. All instructors should follow the same curriculum.	2. The curriculum should be as flexible as possible.
<i>Test Techniques Used in the Exams</i>	
3. Multiple-choice test technique is practical to test our students' language performance.	3. Some other techniques should also be used to test our students' language performance.
<i>Reliability</i>	
4. The exams of the testing office are reliable.	4. My own exams will be more reliable.
<i>Differences in Students' Levels</i>	
5. Centrally administered achievement test is a good way to determine the differences in students' language levels.	5. The instructors should prepare their questions taking their students' different language levels into consideration.

Similar 5-point Likert-type Scale from '5=Strongly Agree' to '1=Strongly Disagree' was used in the fourth part of the questionnaires in order to find out the students' and instructors' perceptions of the exams having been prepared by the testing office so far. This part consists of twelve items each of which aims to get idea on one quality of the 'Compulsory English Language Course' exams. The items in the students' questionnaire show parallelism with the ones in instructors' questionnaire (See the Fourth Part in Students Questionnaire, Appendix D). The items and the characteristic of the exam they represent can be grouped as in the following Table:



*Table 3: Items of the Fourth Part of the Instructors' Questionnaire*

Items Related to...	Exams' _____.
1. The questions which are used in the exams match the course objectives.	Content Validity
2. The content of the questions matches the content I teach in the classroom.	
3. Multiple-choice questions match the activity types that I use in the classroom.	
4. The questions represent the topic that I teach in the classroom.	
5. Multiple-choice test technique is <b>successful</b> in assessing my students' success.	Success
6. Multiple-choice test technique is <b>efficient</b> in assessing my students' success.	Efficiency
7. The questions used in the exams are authentic.	Authenticity
8. In the exams language is tested in the way it is taught.	Content Validity
9. The questions are clear enough to understand.	Clearness
10. Design of these exams is appropriate for my students.	Design
11. Test organization is adequate.	Organization
12. I prefer using only multiple-choice test technique to other test techniques.	Question Types

Finally, fourteen objective test techniques are presented in the fifth parts of the questionnaires. In this part, students and instructors are asked how often these techniques are used in their classrooms and which of these techniques can be used in centrally administered achievement tests. Thus, there are two columns trying to find answers to the questions above. In the first column, students and instructors show how frequently these techniques are used in their classrooms by putting a cross (x) into the most appropriate box of the 5-point Likert-type Scale ranging from '5=Always' to '1=Never'. In the second column, they express their ideas on which techniques can be used in the exams by putting a cross (x) for the most appropriate test technique. 5-point Likert Scale of this part ranges from '5=Strongly Agree' to '1=Strongly Disagree'.

Having described all parts of the questionnaires, the researcher aims to explain how the questionnaires were administered in detail in the following sections of this chapter.

## **3.2 ADMINISTERING THE QUESTIONNAIRES**

The questionnaires were administered both to the students and the instructors. In order to avoid any difficulties in gathering the necessary data, they were first piloted and then used in the main study to collect the data needed. With their setting, participants, materials, procedures for data collection and data analysis, *Pilot and Main Study* are explained below in detail.

### **3.2.1 THE PILOT STUDY**

#### **3.2.1.1 Setting**

The pilot study was conducted at Çanakkale Onsekiz Mart University. Both the students and the instructors participated in the study. Instructors' questionnaire was piloted to the instructors of COMU and to the experts in ELT. Students' questionnaire was administered to the students from different faculties and a college. Piloting was done in November-December, 2007.

#### **3.2.1.2 Participants**

Instructors' questionnaire was piloted to 33 instructors in November, 2007. At that time there were 37 instructors who had worked or were still working as the instructors of 'Compulsory English Language Course'. Thus, nearly 94% of the instructors working at COMU participated in the piloting of the questionnaire. Moreover, the ideas of three experts in 'ELT' and one in 'Evaluation and Measurement' were consulted.

As for the students, 154 questionnaires were handed out. However, 121 students' questionnaires were evaluated. The ones having missing items more than 20% of the total number of the items were not included in the study.

### **3.2.1.3 Materials**

During the piloting process, the researcher handed out two questionnaires – one for the students and one for the instructors. The researcher also attached a piece of paper titled ‘Piloting the Questionnaire’ to the instructors’ questionnaire. Instructors’ ideas were asked with the help of this questionnaire consisting of eight open-ended questions (See Appendix C for ‘Piloting the Questionnaire’). The participants were asked to comment on any issue without hesitation on this piece of paper attached at the end of the questionnaires. Students were also asked to write their comments on the last page of their questionnaires.

Also, the experts checked the questionnaires for any difficulties in understanding the items and for any ambiguous items. All these participants involved and materials used in piloting made it possible for the researcher to make the necessary changes on the questionnaires in order to collect data without any problem in the main study.

### **3.2.1.4 Data Collection**

Data for piloting were collected from 121 volunteer students and 33 instructors. The researcher himself handed out the questionnaires to the students who are available and volunteer for the pilot study. Students were asked to state their ideas on the questionnaire while collecting the data. As for the data collection process of the instructors, the researcher sent them the questionnaires via e-mails. Also, the copies of the instructors’ questionnaire were left to everywhere where instructors could reach easily so that all the instructors could participate in the study. The process for data collection nearly took three weeks (November-December, 2007). The instructors were reminded of the questionnaire via SMS on every fourth day in the data collection process. Then, the data collected from the students and the instructors were analysed to get some implications for the main study.

### 3.2.1.5 Data Analysis

Most of the data obtained from the pilot study were not analysed. Most parts of the questionnaires were not included in the analysis. The instructors', students' and experts' ideas on these parts were evaluated and necessary changes were made in these parts. Their answers to the open-ended questions helped the researcher to construct the ultimate version of the questionnaires for the main study.

According to the instructors' answers, questionnaires took between 10-15 minutes. Two of the instructors wanted the questionnaire to be in Turkish. However, as most of the instructors did not object to the questionnaires' language, it was administered in English in the main study too.

As the twelve items in the fourth parts of the questionnaires question the students' and the instructors' ideas on the centrally administered achievement tests' characteristics, the researcher analysed *the internal consistency* of the items of this part for each questionnaire. Table 4 presents the internal consistency reliability of the items of the fourth parts of the questionnaires.

**Table 4:** *Internal Consistency Reliability (Cronbach Alpha Coefficient) for the Fourth Parts of the Questionnaires in Piloting.*

Questionnaire	Alpha Reliability	n		
		Valid	Excluded	Total
Students' Questionnaire	.81	116	5	121
Instructors' Questionnaire	.91	33	0	33

Twelve items given in the fourth parts of the questionnaires are shown to have a high degree of internal consistency with values *.81* for students' and *.91* for the instructors' questionnaires. These values are acceptable according to Büyüköztürk (2006) who recommends levels of *.70* or greater for scales like these.

### **3.2.1.6 Implications for the Main Study**

Pilot study helped the researcher to make necessary changes in the questionnaires for the main study. Some of the instructors and the experts believed that the introduction parts of the questionnaires had been very long. Thus, the researcher revised the introduction parts for the main study which are shorter and much clearer for the students and the instructors.

As the students who do not want to answer or who do not want to spend time on answering the questionnaires put a cross (x) for the 'No Idea' box, the researcher changed the 'No Idea' section into 'Undecided'. Giving three points for the participants who have no idea about the items of the questionnaires increased the mean values of the results. Therefore, with the advices of the experts in ELT, using 'Undecided' instead of 'No Idea' was found to be more useful for the main study.

In the pilot study, there were headings for each part of the questionnaires. In order to shorten the length of the questionnaires, those headings were omitted. Both questionnaires were shortened about one page for each. Some misspelled and unnecessary words were found during the piloting. Researcher corrected the misspelled words and omitted the unnecessary ones for the main study.

All in all, students', instructors' and some experts' ideas and comments let the researcher revise some of the items and make necessary changes for the main study.

## **3.2.2 MAIN STUDY**

### **3.2.2.1 Setting**

The main study was conducted at COMU. Both the students and the instructors involved in the study. Instructors' questionnaire was administered to the

instructors of COMU. Students' questionnaire was conducted at different faculties and at a college. Main Study was carried out in February-March, 2008.

### 3.2.2.2 Sampling and the Participants

As this study aims to have an idea on the students' and the instructors' perceptions of the 'Compulsory English Language Course' exams at COMU, target population of the study consists of 34 instructors and 2928 second year students. The number of the students was taken from the Students' Affairs Office and the number of the instructors was taken from the Foreign Languages Department on the 13<sup>th</sup> of February, 2008 (see the number of the target population of the students in terms of their genders and departments in Table 11, on page 65)

Instructors' questionnaire was administered to 33 instructors in February, 2008. At that time, there were 34 instructors who had worked or were still working as the instructors of 'Compulsory English Language Course'. Thus, nearly 97% of the instructors working at COMU involved in the main study.

Some information about the instructors who involved in the study will be presented in the following tables. These tables show the numbers of the instructors according to their genders, experiences, educational backgrounds, and graduation departments.

**Table 5: Gender Distribution of the Instructors Involving in the Study**

	<b>Frequency</b>	<b>Percent</b>	<b>Valid Percent</b>	<b>Cumulative Percent</b>
Male	15	45.5	45.5	45.5
Female	18	54.5	54.5	100.0
<b>Total</b>	33	100.0	100.0	

There are 15 males and 18 females involving in the study. As it can be seen in the table above, the number of females working as an instructor is more than that of males.

**Table 6: Experience of the Instructors Involving in the Study**

	<b>Frequency</b>	<b>Percent</b>	<b>Valid Percent</b>	<b>Cumulative Percent</b>
0-3 years	8	24.2	24.2	24.2
4-6 years	15	45.5	45.5	69.7
7 - more years	10	30.3	30.3	100.0
<b>Total</b>	33	100.0	100.0	

As for their teaching experience, there are 8 instructors who are working in their first years. 15 instructors have been working for 4-6 years and seem to be more experienced than the former ones. Finally, there are 10 instructors who have been working for 7 and more years.

**Table 7: Educational Background of the Instructors Involving in the Study**

<b>Graduation</b>	<b>Frequency</b>	<b>Percent</b>	<b>Valid Percent</b>	<b>Cumulative Percent</b>
BA Degree	15	45.5	45.5	45.5
MA Degree	18	54.5	54.5	100.0
<b>Total</b>	33	100.0	100.0	

As it can be seen in the table above, more than half of the instructors have MA degree.

**Table 8: Instructors' BA Degree Graduation Departments**

<b>BA Graduation</b>	<b>Frequency</b>	<b>Percent</b>	<b>Valid Percent</b>	<b>Cumulative Percent</b>
English Language Teaching	23	69.7	69.7	69.7
English Language Literature	6	18.2	18.2	87.9
English Linguistics	1	3.0	3.0	90.9
Translation and Interpretation	1	3.0	3.0	93.9
American Culture and Literature	2	6.1	6.1	100.0
<b>Total</b>	33	100.0	100.0	

Most of the instructors graduated from English Language Teaching (ELT) Departments. They all had 'Testing and Evaluation' courses when they were students in ELT departments. 10 instructors graduated from different departments such as English Language Literature (6), American Culture and Literature (2), English Linguistics (1), and Translation and Interpretation (1).

*Table 9: Instructors' MA Degree Graduation Departments*

<b>MA_Graduation</b>	<b>Frequency</b>	<b>Percent</b>	<b>Valid Percent</b>	<b>Cumulative Percent</b>
English Language Teaching	14	77.8	77.8	77.8
English Language Literature	3	16.7	16.7	94.4
English Linguistics	1	5.6	5.6	100.0
<b>Total</b>	18	100.0	100.0	

Most of the instructors have MA degrees from English Language Teaching (ELT) Departments. There are not any instructors who have MA degrees from 'Translation and Interpretation' or 'American Culture and Literature'. Furthermore, in 'the other' section, one instructor specified that he has his MA degree from Business Administration Department.

*Table 10: Instructors' Background Knowledge on Testing and Evaluation*

	<b>BA Course</b>	<b>MA Course</b>	<b>Seminar</b>	<b>Conference</b>	<b>Symposium</b>
<b>Attended</b>	24	17	11	9	5
<b>Not Attended</b>	9	16	22	24	28
<b>Total</b>	33	33	33	33	33

As it can be seen in the table above, most of the instructors attended a BA Degree Course on 'Testing and Evaluation'. There are some instructors who did not attend any BA courses on 'Testing and Evaluation'. However, they had Testing and Evaluation courses when they were having MA degrees in ELT departments. Also, some instructors stated that they have participated in 'In-service Training Programmes', 'Workshops', and 'Courses by British Council'. All the instructors have participated into at least one of the activities given above. That is to say, they are knowledgeable enough to answer the questionnaire of this study efficiently.

In addition to the facts above, all the instructors who involved in this study have taught either at a faculty or at a college. This means that all of them have enough experience about the testing system being discussed in this study.



As for the students, they are all in their second years at the university. They all had the “Compulsory English Language Course” the year before the study was conducted. Therefore, they are assumed to answer the questions in the questionnaire without any fear of their instructors or the possibility of the changes in the testing system.

The researcher got the total population of the students (2928) from the Students’ Affairs Office. As it is almost impossible for the researcher to reach the whole population of the students (2928), the researcher made a *Stratified Random Sampling*. As suggested by Anderson (1990; cited in Balcı, 2005:95), with the ‘*Confidence Level*’ of 95% and with the ‘*Confidence Interval*’ of 5%; in a population consisting of 5000 members, 356 of these members can be used as a sample while carrying out a research. Thus, the researcher used 0.125 of the total population as a sample of this study which is 367. Stratified Sampling for the students can be seen in the table below.

**Table 11:** *Stratified Random Sampling for Students*

<b>Faculty or College</b>	<b>Male</b>	<b>0.125 of Male Population</b>	<b>Female</b>	<b>0.125 of Female Population</b>	<b>Total</b>	<b>0.125 of the Total Population</b>
Biga Faculty of Economics and Administrative Sciences	359	45	395	49	754	94
Faculty of Education	283	35	401	50	684	85
Faculty of Sciences and Literature	514	64	390	49	904	113
Faculty of Theology	12	2	15	2	27	4
Faculty of Fine Arts	43	6	65	8	108	14
Faculty of Engineering and Architecture	152	19	53	7	205	26
Faculty of Fisheries	43	5	5	1	48	6
Faculty of Agriculture	83	10	21	3	104	13
College Of Health	28	4	66	8	94	12
<b>Total</b>	<b>1517</b>	<b>190</b>	<b>1411</b>	<b>177</b>	<b>2928</b>	<b>367</b>

The reason for choosing the faculties and the college listed above is that they all have four-year education in their programmes and they have common Compulsory English Language Course exams.

### **3.2.2.3 Materials**

Two questionnaires constructed by the researcher were used in the study. They were constructed in order to have an idea on the students' and the instructors' perceptions of the Compulsory English Language Course exams prepared by the testers of the testing office at Çanakkale Onsekiz Mart University. Students' and instructors' questionnaires both helped the researcher to learn about the participants' ideas on the practices of the testing office.

### **3.2.2.4 Data Collection**

Before conducting the main study, the researcher got the list of the instructors, who had taught at a 'Compulsory English Language', from the Foreign Languages Department. In this list, there were 34 instructors to be reached and 33 of them participated in the study. The researcher sent them the questionnaires via e-mails. Moreover, the copies of the instructors' questionnaire were left to everywhere where instructors could reach easily so that all the instructors could participate in the study. The researcher put ticks to their names in the list whenever he got the participants' questionnaires. He sent SMS to those who did not fill in the questionnaire every fourth day in the data collection process. Data collection process from the instructors nearly took 20 days (13 February – 3 March, 2008).

As for the data collection process from the students, the researcher aimed to conduct the study to 367 students. Thus, he made 800 copies of the students' questionnaire. In the first and second weeks of the spring term (18-29 February, 2008), the researcher prepared the questionnaires which were put into separate envelope for each faculty or college. Then, he conducted the questionnaires with the help of the lectures working in those faculties or college in the third and fourth weeks of the spring term (3-14 March, 2008). Data collection process from the students nearly took a month (18 February – 14 March, 2008).

### 3.2.2.5 Data Analysis

The data obtained through the questionnaire were analyzed via “Descriptive Statistics”, “One-way ANOVA”, “Independent Samples T-Test” and “Nonparametric Kruskal-Wallis Test” by using SPSS (Statistical Package for Social Sciences) 13.0 for Windows. The results of these analyses will be presented in the following chapter named ‘*Findings and Discussion*’.

Moreover, the researcher analysed *the internal consistency* of the items of the fourth part for each questionnaire as he did in the pilot study. Following table presents the internal consistency reliability of the items of the fourth parts of the questionnaires.

**Table 12:** *Internal Consistency Reliability (Cronbach Alpha Coefficient) for the Fourth Parts of the Questionnaires in the Main Study*

Questionnaire	Alpha Reliability	n		
		Valid	Excluded	Total
Students' Questionnaire	.84	349	18	367
Instructors' Questionnaire	.86	32	1	33

Twelve items given in the fourth parts of the questionnaires are shown to have a high degree of internal consistency with values .84 for students' and .86 for the instructors' questionnaires. As in the piloting, these values are acceptable for the data to be analyzed.

### 3.3 SUMMARY

This chapter presents the methodology followed in the study. Firstly, the construction process of the questionnaires is explained in detail. Then, a detailed description of the questionnaires is provided. Secondly, how the questionnaires were administered is described. Eventually, the pilot and the main study with their setting, participants, materials, and procedures for data collection and analysis are highlighted in this chapter.

## CHAPTER FOUR

### FINDINGS AND DISCUSSION

#### 4.0 INTRODUCTION

This chapter presents the findings of the statistical analysis of the data obtained from the main study. The findings are presented below the Research Questions (**RQ**). Discussions are made for the research questions below the tables in which the results are shown.

#### 4.1 RESULTS AND DISCUSSION

**RQ1** Are there any differences between the activities that are used by the instructors and the activities that the students expect to be used in their classrooms in order to develop different language skills and areas?

**Table 13:** Results of the Kruskal-Wallis Test Regarding Difference between the Activities that are Used by the **Instructors**, and the Activities that the **Students** Expect to be Used to Develop Language Skills and Areas

Skills and Areas		N	Mean	Std. Dev. (SD)	df	Chi-Square (X <sup>2</sup> )	p	Significant Difference
Teaching Listening	Students	367	3.8937	.97630	1	10.454	.001*	A- B**
	Instructors	33	3.2121	1.19262				
Teaching Speaking	Students	367	4.3433	.91547	1	18.575	.000*	A- B**
	Instructors	33	3.6970	.91804				
Teaching Reading	Students	367	3.9155	.90291	1	1.193	.275	-----
	Instructors	33	4.1212	.78093				
Teaching Writing	Students	366	3.6366	1.04511	1	7.686	.006*	A- B**
	Instructors	33	3.1818	.80834				
Teaching Grammar	Students	366	4.0082	.97218	1	15.058	.000*	A- B**
	Instructors	33	4.6061	.82687				
Teaching Vocabulary	Students	367	4.4550	.79102	1	.547	.459	-----
	Instructors	33	4.5758	.66287				
Teaching Pronunciation	Students	367	4.3706	.87091	1	28.162	.000*	A- B**
	Instructors	33	3.5758	.86712				

\* p < .05

\*\* A= Students' Ideas, B= Instructors' Ideas

In order to find the difference between the activities that are used by the instructors and the activities that the students expect to be used in their classrooms to develop different language skills and areas, Nonparametric Kruskal-Wallis Test was done as the number of the participants was limited. The results can be seen in the table above. According to the results, there are not any significant differences in terms of reading and vocabulary. Namely, the instructors use reading and vocabulary activities in the nearly same frequency that the students expect them to be used. However, there are significant differences in pronunciation [ $X^2_{(1)} = 28.162$ ,  $p < .05$ ], speaking [ $X^2_{(1)} = 18.575$ ,  $p < .05$ ], listening [ $X^2_{(1)} = 10.454$ ,  $p < .05$ ] and writing [ $X^2_{(1)} = 7.686$ ,  $p < .05$ ] as students expect these skills and areas to be used more frequently than their instructors use them in their classes. Furthermore, there is also significant difference in terms of the use of grammar [ $X^2_{(1)} = 15.058$ ,  $p < .05$ ] in the class. As for grammar, instructors use grammar more than the students expect it to be used [ $\bar{X}_{(Instructors)} = 4.6061$ ,  $\bar{X}_{(Students)} = 4.0082$ ]. In the light of these findings, the instructors should focus on the activities improving students speaking, listening and writing skills. They must also do some pronunciation practices. Students seem to need more communicative lessons. Focusing on communicative lessons means communicative exams. In addition to grammar, the exams should include other language skills and areas as well.

**RQ1-A** Which language skill or area do the students expect to be developed most in their lessons?

As it can be seen in Table 13, students respectively want to improve their vocabulary [ $\bar{X}_{(Vocabulary)} = 4.4550$ ], pronunciation [ $\bar{X}_{(Pronunciation)} = 4.3706$ ] and speaking skills [ $\bar{X}_{(Speaking)} = 4.3433$ ] most.

**RQ1-B** Which language skill or area do the instructors try to develop most in their lessons?

Instructors respectively want to improve their students' grammar [ $\bar{X}_{(Grammar)} = 4.6061$ ], vocabulary [ $\bar{X}_{(Vocabulary)} = 4.5758$ ] and reading skills [ $\bar{X}_{(Reading)} = 4.1212$ ] most.

The results of **RQ1-A** and **RQ1-B** show us that the instructors should also attach necessary importance to speaking and pronunciation. Teaching communicative aspects of language is necessary in foreign language learning classes. The standardized tests applied at COMU test only grammar, vocabulary and reading comprehension. The ways to integrate all skills and areas both to our lessons and to our exams should be found and implemented. Thus, different test techniques assessing different language skills and areas should be used.

**RQ2** Do the students and the instructors prefer the Compulsory English Language Course exams to be prepared by the testing office or by the instructor of the course?

**Table 14:** Instructors' Ideas on Whom to Organize the Exams

Who Should Organize the Exams?	Frequency	Percent	Valid Percent	Cumulative Percent
Testing Office	24	72.7	72.7	72.7
Instructor	9	27.3	27.3	100.0
Total	33	100.0	100.0	

In Table 14, it can be seen that nearly 73 % of the instructors want the exam to be prepared by a testing office. Only nine of the instructors (nearly 27% of the total population) prefer preparing the exam by themselves. Most of the instructors seem to be satisfied with the testing office according to these results or have some other reasons to prefer testing office.

**Table 15:** Students' Ideas on Whom to Organize the Exams

Who Should Organize the Exams?	Frequency	Percent	Valid Percent	Cumulative Percent
Testing Office	126	34.3	34.3	34.3
Instructor	241	65.7	65.7	100.0
Total	367	100.0	100.0	

Unlike the instructors, more than half of (65.7%) the students want the exam to be prepared by the instructor of the course. 34.3 percent of the students prefer testing office for the organization process of the Compulsory English Language Course Exams. While instructors prefer testing office, students prefer the instructor of the course for the organization of the exams. Their most common reasons for preferring either testing office or the instructor of the course will be discussed below.

**RQ2-A** What is their most common reason for preferring either testing office or the instructor for the preparation and organization process of the exam?

**Table 16:** The Instructors' Reasons for Preferring the Testing Office

<b>Reasons for Preferring Testing Office</b>	<b>N</b>	<b>Mean</b>	<b>Min.</b>	<b>Max.</b>	<b>SD</b>
1. All the instructors should follow the same language teaching materials.	24	4.2500	2.00	5.00	.94490
2. All instructors should follow the same curriculum.	24	4.6667	3.00	5.00	.56466
3. Multiple-choice test technique is practical to test our students' language performance.	24	3.9583	2.00	5.00	.95458
4. The exams of the testing office are reliable.	24	4.0000	2.00	5.00	.93250
5. Centrally administered achievement test is a good way to determine the differences in students' language levels.	24	3.9583	2.00	5.00	.80645

Instructors' reasons for preferring the testing office to organize the exams are presented in the table above. Among the 24 instructors who believe that testing office should organize the exams, there is nobody who strongly disagrees (= '1.00') with the reasons specified in the questionnaire. When the mean values of the items are taken into consideration, the results show that they all agree with the reasons listed in the questionnaire. However, their most common reason to prefer testing office is that 'All instructors should follow the same curriculum' ( $\bar{X} = 4.67$ ). They believe that if the exams were not prepared by the testing office, the instructors would not follow the curriculum.

**Table 17:** The Instructors' Reasons for Preferring the Instructor of the Course

<b>Reasons for Preferring Instructor of the Course</b>	<b>N</b>	<b>Mean</b>	<b>Min.</b>	<b>Max.</b>	<b>SD</b>
1. I want to use extra language teaching materials and prepare my own exam accordingly.	9	4.7778	4.00	5.00	.44096
2. The curriculum should be as flexible as possible.	9	4.1111	2.00	5.00	.92796
3. Some other techniques should also be used to test our students' language performance.	9	4.5556	3.00	5.00	.72648
4. My own exams will be more reliable.	9	4.5556	3.00	5.00	.72648
5. The instructors should prepare their questions by taking their students' different language levels into consideration.	9	4.6667	4.00	5.00	.50000

Moreover, among the 9 instructors who believe that the instructor of the

course should organize the exams, there is nobody who strongly disagrees (= '1') with the reasons specified in the questionnaire. Instructors' reasons for preferring instructor of the course to organize the exams are presented in Table 17.

The results again show that they all agree with the reasons listed in the questionnaire. The most common reason for the instructors' preference is that they want to use extra language teaching materials and prepare their own exam accordingly ( $\bar{X} = 4.78$ ). The least common item is that 'the curriculum should be as flexible as possible' ( $\bar{X} = 4.11$ ). In the light of these findings, it seems that following a common curriculum is very important for the instructors.

**Table 18:** The Students' Reasons for Preferring the Testing Office to Organize the Exams

<b>Reasons for Preferring Testing Office</b>	<b>N</b>	<b>Mean</b>	<b>SD</b>
1. Centrally administered achievement tests determine the differences in students' levels better.	126	4.2778	.84512
2. Multiple-choice test technique is used in the exams.	126	4.1667	.98590
3. Centrally administered achievement tests measure students' performance fairly.	126	4.0952	.92458
4. Centrally administered achievement tests are more reliable.	125	4.3120	.83681
5. Centrally administered achievement tests hinder the instructors' possible negative attitudes towards their learners.	126	4.4444	.89938

Students' reasons for preferring the testing office to organize the exams are presented in the table above. 126 students believe that testing office should organize the exams. Mean values of the items show that they all agree with the reasons listed in the questionnaire. However, their most common reason to prefer testing office is that 'Centrally administered achievement tests hinder the instructors' possible negative attitudes towards their learners.' ( $\bar{X} = 4.44$ ). Students believe that if the exams were not prepared by the testing office, the instructors would reflect their possible negative attitudes to the students' grades or to the exam questions. Although students think that the exams by testing office are fair ( $\bar{X} = 4.10$ ), the third item (Centrally administered achievement tests measure students' performance fairly) is the least common reason for the students' preference.



**Table 19:** The Students' Reasons for Preferring the Instructor of the Course to Organize the Exams

Reasons for Preferring Instructor of the Course	N	Mean	SD
1. The instructors prepare their questions by taking their students' different language levels into consideration.	241	4.6722	.73910
2. Some other test techniques can also be used in the exams.	241	4.0498	1.03964
3. As there are differences in students' levels, it is fairer that the instructors prepare their own exams.	241	4.6017	.71810
4. The exams prepared by the instructor of the course are more reliable.	241	3.5519	1.23761
5. Instructors' possible positive attitudes towards their learners affect the evaluation process positively.	241	4.2158	.98060

As discussed before, more than half of (65.7%) the students want the exam to be prepared by the instructor of the course. Their most common reason for preferring the instructor of the course is that 'the instructors prepare their questions by taking their students' different language levels into consideration' ( $\bar{X} = 4.67$ ). Students want the exam to be prepared according to their levels. Although most of the students want the exams to be prepared by the instructor of the course, they have some doubts about those exams' reliability ( $\bar{X} = 3.55$ ).

**RQ3** What do the instructors and the students think about the exams that have been prepared by the test constructors of the testing office so far?

**Table 20:** The Instructors' Perceptions of the Exams

Instructors' Perceptions of the Exams	N	Mean	SD
1. The questions which are used in the exams match the course objectives.	33	3.8788	.89294
2. The content of the questions matches the content I teach in the classroom.	33	4.0000	.86603
3. Multiple-choice questions match the activity types that I use in the classroom.	33	3.1212	1.11124
4. The questions represent the topic that I teach in the classroom.	33	4.0000	.86603
5. Multiple-choice test technique is <b>successful</b> in assessing my students' success.	33	3.3333	.98953
6. Multiple-choice test technique is <b>efficient</b> in assessing my students' success.	32	3.2188	1.09939
7. The questions used in the exams are authentic.	33	3.1212	1.19262
8. In the exams, language is tested in the way it is taught.	33	2.9091	1.15552
9. The questions are clear enough to understand.	33	4.1818	.52764
10. Design of these exams is appropriate for my students.	33	3.8788	.73983
11. Test organization is adequate.	33	3.5152	1.14895
12. I prefer using only multiple-choice test technique to other test techniques.	33	2.5152	1.25303
<b>Total Mean of the items above:</b>	33	3.4646	

Total mean of the instructors' perceptions ( $\bar{X} = 3.46$ ) shows that they are not much sure about the efficiency of the exams' different characteristics. Although most of the instructors prefer the exam to be prepared by the testing office, they do not

seem to be very satisfied with its current practices. Instructors agree that ‘the questions are clear enough to understand with the highest mean value of 4.18. Furthermore, they believe that the content of the questions matches the content they teach ( $\bar{X} = 4.00$ ) and the questions represent the topic that they teach in the classroom ( $\bar{X} = 4.00$ ). This shows that instructors agree that the exams have content validity. However, they do not agree that language is tested in the way it is taught in the exams ( $\bar{X} = 2.91$ ). This may be because the instructors use many question types while teaching English but only multiple-choice while testing what they have taught. Hence, using other test techniques while assessing our learners’ language performance will help the testers construct better tests. Mean value of the last item ( $\bar{X} = 2.52$ ) confirms this result. That is to say, instructors prefer using other test techniques in addition to the multiple-choice test technique.

**Table 21:** The Students’ Perceptions of the Exams

<b>Students’ Perceptions of the Exams</b>	<b>N</b>	<b>Mean</b>	<b>SD</b>
1. The questions which are used in the exams match the course objectives.	367	3.7139	.90379
2. The content of the questions matches the content I learned in the classroom.	367	3.7193	.94391
3. Multiple-choice questions match the activity types that are used in the classroom.	364	3.5989	.99231
4. The questions represent the topic that I learned in the classroom.	366	3.4071	1.06029
5. Multiple-choice questions are <b>successful</b> in assessing my language performance.	367	3.0000	1.16178
6. Multiple-choice questions are <b>efficient</b> in assessing my language performance.	363	2.8733	1.12502
7. The questions used in the exams are authentic.	365	3.2055	1.00765
8. In the exams, language is tested in the way it is taught.	366	3.4153	1.02151
9. The questions are clear enough to understand.	365	3.5507	1.13911
10. Design of these exams is appropriate for the students.	367	3.5068	1.02127
11. Test organization is adequate.	362	3.3674	1.06067
12. I prefer only multiple-choice test technique rather than other ones in the exams.	367	3.3651	1.32553
<b>Total Mean of the items above:</b>	367	3.3801	

When the total mean of the items above is examined ( $\bar{X} = 3.38$ ), students are undecided about the statements regarding the exams prepared by the testing office so far. Like the instructors, students also agree that the content of the questions matches the content they learned in the classroom ( $\bar{X} = 3.72$ ) and the questions which are used in the exams match the course objectives ( $\bar{X} = 3.71$ ). However, they do not

agree that multiple-choice questions are **efficient** ( $\bar{X} = 2.87$ ) and **successful** ( $\bar{X} = 3.00$ ) in assessing their language performance as these items have the lowest mean values respectively. Therefore, using other test techniques in centrally administered achievement tests is possible according to students' and instructors' ideas.

**RQ3-A** Is there a significant difference among students' thoughts about the exams prepared by the testing office in terms of their success?

Among 367 students, 57 of them failed the compulsory English language course (0-59 Points – unsuccessful). 121 of them were successful enough to pass the course (60-79 points – successful). Finally, 188 of them were very successful in the course (80-100 Points – very successful). In order to find out whether there is a significant difference among students' thoughts about the exams prepared by the testing office in terms of their success, the researcher carried out ANOVA test.

**Table 22:** Results of ANOVA Regarding the Difference between Students' Success and their Perceptions of the Exams

Students' Perceptions of the Exams	Source of Variance	Sum of Squares	df	Mean Square	F	p	Sig. Difference (Scheffe)
1. The questions which are used in the exams match the course objectives.	Between Groups	5.019	2	2.509	3.107	.046**	*1-3
	Within Groups	293.941	364	.808			
	Total	298.959	366				
2. The content of the questions matches the content I learned in the classroom.	Between Groups	6.552	2	3.276	3.732	.025**	*1-3
	Within Groups	319.541	364	.878			
	Total	326.093	366				
3. Multiple-choice questions match the activity types that are used in the class.	Between Groups	5.767	2	2.884	2.960	.053	-----
	Within Groups	351.672	361	.974			
	Total	357.440	363				
4. The questions represent the topic that I learned in the classroom.	Between Groups	13.433	2	6.716	6.143	.002**	*1-2 *1-3
	Within Groups	396.909	363	1.093			
	Total	410.342	365				
5. Multiple-choice is <b>successful</b> in assessing language performance.	Between Groups	5.050	2	2.525	1.880	.154	-----
	Within Groups	488.950	364	1.343			
	Total	494.000	366				
6. Multiple-choice is <b>efficient</b> in assessing my language performance.	Between Groups	5.060	2	2.530	2.010	.135	-----
	Within Groups	453.111	360	1.259			
	Total	458.171	362				
7. The questions used in the exams are authentic.	Between Groups	1.166	2	.583	.573	.564	-----
	Within Groups	368.423	362	1.018			
	Total	369.589	364				

8. In the exams, language is tested in the way it is taught.	Between Groups	.913	2	.456	.436	.647	-----
	Within Groups	379.962	363	1.047			
	Total	380.874	365				
9. The questions are clear enough to understand.	Between Groups	16.184	2	8.092	6.422	.002**	*1-3
	Within Groups	456.128	362	1.260			
	Total	472.312	364				
10. Design of these exams is appropriate for the students.	Between Groups	7.719	2	3.859	3.756	.024**	*1-2 *1-3
	Within Groups	374.014	364	1.028			
	Total	381.733	366				
11. Test organization is adequate.	Between Groups	9.857	2	4.928	4.465	.012**	*1-2
	Within Groups	396.278	359	1.104			
	Total	406.135	361				
12. I prefer only multiple-choice rather than other ones in the exams.	Between Groups	.338	2	.169	.096	.909	-----
	Within Groups	642.736	364	1.766			
	Total	643.074	366				

\*1 = 0-59 points (unsuccessful), 2 = 60-79 points (successful), 3 = 80-100 points (very successful)

\*\* p < .05

When both Table 22 and 23 are examined, the results suggest that there are significant differences among the groups in terms of the six items listed above. Firstly, very successful learners ( $\bar{X}_{(VS)} = 3.78$ ) believe that ‘the questions which are used in the exams match the course objectives’ more than the unsuccessful learners ( $\bar{X}_{(US)} = 3.45$ ) do [ $F_{(2)} = 3.107$ ,  $p < .05$ ]. Similarly, very successful learners ( $\bar{X}_{(VS)} = 3.79$ ) believe that ‘the content of the questions matches the content they learned in the classroom’ more than the unsuccessful learners ( $\bar{X}_{(US)} = 3.41$ ) do [ $F_{(2)} = 3.732$ ,  $p < .05$ ]. Also, as the students are more successful, they think that ‘the questions represent the topic that they learned in the classroom’ more [ $F_{(2)} = 6.143$ ,  $p < .05$ ]. According to the results, it can be discussed that normally the more successful the students are, the more content validity they believe the exams have.

Very successful learners ( $\bar{X}_{(VS)} = 3.71$ ) find the questions more clear than the unsuccessful ones ( $\bar{X}_{(US)} = 3.10$ ) [ $F_{(2)} = 6.422$ ,  $p < .05$ ]. Moreover, successful learners ( $\bar{X}_{(S)} = 3.58$ ) think that ‘test organization is adequate’ more than the unsuccessful learners ( $\bar{X}_{(US)} = 3.12$ ) do [ $F_{(2)} = 4.465$ ,  $p < .05$ ]. Finally, the more successful the students are, the more appropriate they find the design of the exams [ $F_{(2)} = 3.756$ ,  $p < .05$ ].

**Table 23:** Mean Values of the Unsuccessful, Successful and Very Successful Students' Ideas

Students' Perceptions of the Exams	Students' Success	N	Mean	Std. Deviation
* 1. The questions which are used in the exams match the course objectives.	0-59 Points	58	3.4483	.99424
	60-79 Points	121	3.7355	.95541
	80-100 Points	188	3.7819	.82753
* 2. The content of the questions matches the content I learned in the classroom.	0-59 Points	58	3.4138	.99180
	60-79 Points	121	3.7521	.93346
	80-100 Points	188	3.7926	.92184
3. Multiple-choice questions match the activity types that are used in the class.	0-59 Points	57	3.3333	1.12335
	60-79 Points	121	3.7190	1.02651
	80-100 Points	186	3.6022	.91403
* 4. The questions represent the topic that I learned in the classroom.	0-59 Points	57	2.9649	1.10138
	60-79 Points	121	3.4545	1.05672
	80-100 Points	188	3.5106	1.02111
5. Multiple-choice is <b>successful</b> in assessing language performance.	0-59 Points	58	2.8793	1.20055
	60-79 Points	121	3.1653	1.06729
	80-100 Points	188	2.9309	1.20183
6. Multiple-choice is <b>efficient</b> in assessing my language performance.	0-59 Points	57	2.8070	1.18681
	60-79 Points	119	3.0420	1.12291
	80-100 Points	187	2.7861	1.10094
7. The questions used in the exams are authentic.	0-59 Points	57	3.2105	.90113
	60-79 Points	121	3.2810	1.00186
	80-100 Points	187	3.1551	1.04342
8. In the exams language is tested in the way it is taught.	0-59 Points	58	3.3103	1.01233
	60-79 Points	121	3.4628	1.00864
	80-100 Points	187	3.4171	1.03556
* 9. The questions are clear enough to understand.	0-59 Points	58	3.1034	1.16513
	60-79 Points	120	3.5250	1.22277
	80-100 Points	187	3.7059	1.03908
* 10. Design of these exams is appropriate for the students.	0-59 Points	58	3.1724	1.01113
	60-79 Points	121	3.5785	1.04684
	80-100 Points	188	3.5638	.99258
* 11. Test organization is adequate.	0-59 Points	58	3.1207	1.07732
	60-79 Points	120	3.5833	1.07362
	80-100 Points	184	3.3043	1.02684
12. I prefer only multiple-choice rather than other ones in the exams.	0-59 Points	58	3.4310	1.27199
	60-79 Points	121	3.3388	1.35126
	80-100 Points	188	3.3617	1.33125

\* Significant difference among the groups was found ( $p < .05$ ).

Mean values of the items where significant differences (\*) were found among the groups display that *the more successful the students are, the more satisfied they are with the exams' different characteristics.*

**RQ3-B** Is there a significant difference among instructors' thoughts about the exams prepared by the testing office in terms of their job experience?

**Table 24:** The Results of Kruskal-Wallis Test Regarding the Difference between Instructors' Job Experience and their Perceptions of the Exams

Instructors' Perceptions of the Exams	Instructors' Experience	N	Mean Rank	df	Chi-Square $X^2$	p
1. The questions which are used in the exams match the course objectives.	0-3 years	8	13.19	2	2.172	.338
	4-6 years	15	19.10			
	7 – more	10	16.90			
2. The content of the questions matches the content I teach in the classroom.	0-3 years	8	15.50	2	.335	.846
	4-6 years	15	17.70			
	7 – more	10	17.15			
3. Multiple-choice questions match the activity types that I use in the classroom.	0-3 years	8	14.94	2	.973	.615
	4-6 years	15	16.63			
	7 – more	10	19.20			
4. The questions represent the topic that I teach in the classroom.	0-3 years	8	17.19	2	.760	.684
	4-6 years	15	15.70			
	7 – more	10	18.80			
5. Multiple-choice test technique is <b>successful</b> in assessing my students' success.	0-3 years	8	13.38	2	4.775	.092
	4-6 years	15	15.50			
	7 – more	10	22.15			
<b>* 6. Multiple-choice test technique is <i>efficient</i> in assessing my students' success.</b>	<b>0-3 years</b>	<b>8</b>	<b>13.56</b>	2	<b>7.276</b>	<b>.026</b>
	<b>4-6 years</b>	<b>14</b>	<b>13.61</b>			
	<b>7 – more</b>	<b>10</b>	<b>22.90</b>			
7. The questions used in the exams are authentic.	0-3 years	8	16.00	2	.596	.742
	4-6 years	15	16.27			
	7 – more	10	18.90			
8. In the exams, language is tested in the way it is taught.	0-3 years	8	16.69	2	.360	.835
	4-6 years	15	16.20			
	7 – more	10	18.45			
9. The questions are clear enough to understand.	0-3 years	8	15.94	2	.910	.634
	4-6 years	15	16.27			
	7 – more	10	18.95			
10. Design of these exams is appropriate for my students.	0-3 years	8	14.31	2	1.396	.497
	4-6 years	15	17.13			
	7 – more	10	18.95			
11. Test organization is adequate.	0-3 years	8	12.63	2	2.343	.310
	4-6 years	15	18.63			
	7 – more	10	18.05			
12. I prefer using only multiple-choice test technique to other test techniques.	0-3 years	8	12.75	2	2.457	.293
	4-6 years	15	17.53			
	7 – more	10	19.60			

\* Significant difference among the groups was found ( $p < .05$ ).

In order to find out whether there is a significant difference between instructors' perceptions of the exams and their job experience, non-parametric Kruskal-Wallis Test was carried out. Among the twelve items evaluated, only the *sixth item* proved to have a significant difference between instructors' perceptions of the exams and their job experience [ $X^2_{(2)} = 7.276, p < .05$ ]. It can clearly be seen that as the instructors' job experience increase, the mean rank of the item also increases. That is to say, *the more experienced they are, the more efficient they find the exams prepared by the testers of the testing office*. This can be because they do not want to prepare their own exams and they want to use a SAT.

**RQ4** What are students' and instructors' perceptions of the objective test techniques that are used and that can be used in the Standardized Achievement Tests of Compulsory English Language Course?

To find out the students' and the instructors' perceptions of the objective test techniques, five research questions were asked. Independent Samples T-Test, Descriptive Statistics, and Nonparametric Kruskal-Wallis test help the researcher answer those sub-questions regarding their perceptions of the objective test techniques.

**RQ4-A** Are there any significant differences between the students' and instructors' ideas on how frequently objective test techniques are used in the class?

The answer to this research question shows whether there is a consistency between the test techniques that the instructors state they have used in the class and the ones that the students state their instructors used in the class. Among the 14 test techniques, 10 of them show consistency. Independent Samples Test results in Table 25 points out no significant differences among the instructors' and students' choices for the ten of test techniques listed in the questionnaires. However, there are significant differences between students' and instructors' ideas on the frequency of using Short Answer questions [ $t_{(398)} = 2.50, p < .05$ ], True-False Questions [ $t_{(398)} = 2.20, p < .05$ ], Cloze Test [ $t_{(394)} = 2.15, p < .05$ ] and C-Test [ $t_{(39.76)} = 3.32, p < .05$ ].

**Table 25:** The Results of Independent Samples T-Test Regarding the Difference Between Students' and Instructors' Ideas on How Frequently Objective Test Techniques Are Used in the Class

Objective Test Techniques that are used in the Class	Students' - Instructors' Ideas	N	Mean	Std. Deviation	df	t	p
1. Multiple-Choice Questions	Students'	366	3.3825	1.20097	397	-1.182	.238
	Instructors'	33	3.6364	.92932			
* 2. Short-Answer Questions	Students'	367	3.6104	1.00175	398	-2.499	.013
	Instructors'	33	4.0606	.86384			
* 3. True-False Questions	Students'	367	3.2616	1.20097	398	-2.201	.028
	Instructors'	33	3.7273	1.17985			
4. Matching	Students'	365	3.3397	1.12156	396	-1.751	.081
	Instructors'	33	3.6970	1.13150			
5. Completion	Students'	367	3.4278	1.18948	398	-1.254	.210
	Instructors'	33	3.6970	1.07485			
* 6. Cloze Test	Students'	363	2.8650	1.20127	394	2.154	.032
	Instructors'	33	2.3939	1.22320			
* 7. C-Test	Students'	360	2.2333	1.15438	39.763	3.316	.002
	Instructors'	33	1.6061	1.02894			
8. Cloze Elide Test	Students'	363	2.1047	1.15153	394	1.078	.282
	Instructors'	33	1.8788	1.16613			
9. Ordering Tasks (Rearrangement)	Students'	366	3.1803	1.26718	41.553	.795	.431
	Instructors'	33	3.0303	1.01504			
10. Error Correction	Students'	363	3.0882	1.29723	394	-.141	.888
	Instructors'	33	3.1212	1.24392			
11. Transformation	Students'	359	2.7660	1.29712	43.227	1.540	.131
	Instructors'	33	2.4848	.97215			
12. Combination	Students'	367	2.9564	1.24719	398	1.846	.066
	Instructors'	33	2.5455	.93845			
13. Addition	Students'	366	3.1639	1.23883	41.399	.883	.383
	Instructors'	33	3.0000	1.00000			
14. Word Changing	Students'	367	3.4087	1.29384	397	.671	.503
	Instructors'	32	3.2500	1.16398			

\* Significant difference between the groups was found ( $p < .05$ ).

When the mean values are taken into consideration, instructors believe that they use short-answer [ $\bar{X}_{(Ins)} = 4.06$ ;  $\bar{X}_{(Std)} = 3.61$ ] and true-false [ $\bar{X}_{(Ins)} = 3.73$ ;  $\bar{X}_{(Std)} = 3.26$ ] questions more frequently than students believe that their instructors use in the class. The difference can result in students' not taking the activity types in their books into their consideration as the book consists of many short-answer and true-false questions.



The difference for Cloze Test and C-Test is completely different from the other two techniques. Students believe that their instructors use Cloze Test [ $\bar{X}_{(Std)} = 2.87$ ;  $\bar{X}_{(Ins)} = 2.39$ ] and C-Test [ $\bar{X}_{(Std)} = 2.23$ ;  $\bar{X}_{(Ins)} = 1.61$ ] more frequently than their instructors state that they have used in the class. This may be because students did not read the explanations of these techniques and most of them do not know anything about these two techniques.

**RQ4-B** Which objective test techniques do the instructors use most and least frequently in the classroom activities of Compulsory English Language Courses?

As clearly seen in Table 25, short answer questions ( $\bar{X} = 4.06$ ), true-false questions ( $\bar{X} = 3.73$ ), matching and completion items ( $\bar{X} = 3.70$ ) are used by the instructors more frequently than the other test techniques. Multiple-choice test technique which is currently the only way to assess our learners' language performance is the fifth frequently used technique by the instructors ( $\bar{X} = 3.64$ ). C-Test ( $\bar{X} = 1.61$ ), Cloze Elide Tests ( $\bar{X} = 1.88$ ), and Cloze Tests ( $\bar{X} = 2.39$ ) are among the least frequently test techniques used by the instructors. This can be because they do not know much about these techniques. As they are used more frequently than the multiple-choice test technique in classroom activities; short answer questions, true-false questions, matching and completion items can also be used in the Standardized Achievement Tests of COMU.

**RQ4-C** What other objective test techniques can be used in the mid-term and final exams of the course in the future?

As it can be seen in Table 26; according to the instructors, multiple-choice ( $\bar{X} = 4.36$ ), matching ( $\bar{X} = 4.06$ ), ordering tasks ( $\bar{X} = 4.03$ ), completion ( $\bar{X} = 3.91$ ), true false ( $\bar{X} = 3.91$ ), short answer ( $\bar{X} = 3.79$ ), error correction ( $\bar{X} = 3.76$ ) and word changing ( $\bar{X} = 3.67$ ) are among the most preferred objective test techniques. Instructors are undecided about using other techniques in the standardized achievement tests of Compulsory English Language Courses.

As for the students; multiple-choice ( $\bar{X} = 4.31$ ), short answer ( $\bar{X} = 3.97$ ), true false ( $\bar{X} = 3.92$ ), completion ( $\bar{X} = 3.86$ ), matching ( $\bar{X} = 3.78$ ), ordering tasks ( $\bar{X} = 3.72$ ), word changing ( $\bar{X} = 3.72$ ) and error correction ( $\bar{X} = 3.59$ ) are among the most preferred objective test techniques by the students. Students are also undecided about using other techniques in the standardized achievement tests as they are not much familiar with these techniques.

**Table 26:** Students' and Instructors' Perceptions of the Objective Test Techniques that can be used in the Standardized Exams

Objective Test Techniques that can be used in the Standardized Exams	Students' - Instructors' Ideas	N	Mean	Std. Deviation
1. Multiple-Choice Questions	Students'	366	4.3115	.76300
	Instructors'	33	4.3636	.65279
2. Short-Answer Questions	Students'	367	3.9700	1.01178
	Instructors'	33	3.7879	.96039
3. True-False Questions	Students'	367	3.9155	.97563
	Instructors'	33	3.9091	1.04174
4. Matching	Students'	365	3.7836	1.06118
	Instructors'	33	4.0606	.99810
5. Completion	Students'	367	3.8583	1.14833
	Instructors'	33	3.9091	.94748
6. Cloze Test	Students'	362	3.3508	1.25485
	Instructors'	32	3.2813	1.22433
7. C-Test	Students'	359	3.0084	1.20865
	Instructors'	33	2.7879	1.02340
8. Cloze Elide Test	Students'	363	2.9036	1.20040
	Instructors'	33	3.0000	1.11803
9. Ordering Tasks (Rearrangement)	Students'	366	3.7213	1.13665
	Instructors'	32	4.0313	.93272
10. Error Correction	Students'	364	3.5934	1.19897
	Instructors'	33	3.7576	1.11888
11. Transformation	Students'	359	3.1671	1.20053
	Instructors'	33	2.9697	.80951
12. Combination	Students'	367	3.4033	1.22175
	Instructors'	33	3.3030	.88335
13. Addition	Students'	366	3.5738	1.15570
	Instructors'	33	3.4545	.93845
14. Word Changing	Students'	367	3.7193	1.18055
	Instructors'	33	3.6667	1.13652

It is a surprisingly important finding that although their order of preferring objective test techniques slightly differs, *the first eight test techniques that the students and instructors prefer to be used in the exams are completely same.* Thus, these techniques should also be used at SATs.

**RQ4-D** Do the objective test techniques that are used by the instructors in their classroom activities differ in terms of instructors' educational backgrounds?

**Table 27:** *The Results of Kruskal-Wallis Test Regarding the Difference between Instructors' Educational Backgrounds and the Objective Test Techniques They Use*

Objective Test Techniques that are used in the Class	Graduation Degree	N	Mean Rank	df	Chi-Square X <sup>2</sup>	p
1. Multiple-Choice Questions	BA	15	19.20	1	1.575	.210
	MA	18	15.17			
2. Short-Answer Questions	BA	15	19.53	1	2.179	.140
	MA	18	14.89			
3. True-False Questions	BA	15	16.07	1	.285	.594
	MA	18	17.78			
4. Matching	BA	15	18.07	1	.368	.544
	MA	18	16.11			
5. Completion	BA	15	17.90	1	.257	.612
	MA	18	16.25			
6. Cloze Test	BA	15	14.77	1	1.598	.206
	MA	18	18.86			
7. C-Test	BA	15	13.97	1	3.608	.057
	MA	18	19.53			
8. Cloze Elide Test	BA	15	15.00	1	1.421	.233
	MA	18	18.67			
9. Ordering Tasks (Rearrangement)	BA	15	15.07	1	1.210	.271
	MA	18	18.61			
10. Error Correction	BA	15	16.20	1	.217	.642
	MA	18	17.67			
11. Transformation	BA	15	17.47	1	.070	.791
	MA	18	16.61			
12. Combination	BA	15	16.33	1	.144	.704
	MA	18	17.56			
13. Addition	BA	15	17.67	1	.154	.694
	MA	18	16.44			
14. Word Changing	BA	15	16.10	1	.059	.808
	MA	17	16.85			

According to the results of Kruskal-Wallis Test (Table 27), there occur no significant differences between instructors' educational backgrounds and the objective test techniques that they use in their classrooms.

**RQ4-E** Do the instructors' ideas on which objective test techniques can be used in Centrally Administered Achievement tests differ in terms of their educational backgrounds?

**Table 28:** *The Results of Kruskal-Wallis Test Regarding the Difference between Instructors' Educational Backgrounds and Their Ideas on the Objective Test Techniques that can be Used in the Exams*

Objective Test Techniques that are used in the Class	Graduation Degree	N	Mean Rank	df	Chi-Square X <sup>2</sup>	p
1. Multiple-Choice Questions	BA	15	19.03	1	1.596	.207
	MA	18	15.31			
2. Short-Answer Questions	BA	15	20.03	1	3.120	.077
	MA	18	14.47			
3. True-False Questions	BA	15	15.70	1	.567	.451
	MA	18	18.08			
4. Matching	BA	15	15.47	1	.837	.360
	MA	18	18.28			
5. Completion	BA	15	16.53	1	.073	.787
	MA	18	17.39			
6. Cloze Test	BA	15	18.40	1	1.234	.267
	MA	17	14.82			
7. C-Test	BA	15	16.40	1	.117	.733
	MA	18	17.50			
8. Cloze Elide Test	BA	15	15.73	1	.505	.477
	MA	18	18.06			
9. Ordering Tasks (Rearrangement)	BA	15	16.83	1	.042	.837
	MA	17	16.21			
10. Error Correction	BA	15	17.20	1	.013	.910
	MA	18	16.83			
11. Transformation	BA	15	17.90	1	.273	.601
	MA	18	16.25			
12. Combination	BA	15	15.93	1	.386	.534
	MA	18	17.89			
13. Addition	BA	15	17.83	1	.244	.621
	MA	18	16.31			
14. Word Changing	BA	15	17.10	1	.003	.955
	MA	18	16.92			

The results of Kruskal-Wallis Test (Table 28) show that there are no significant differences between Instructors' Educational Backgrounds and their ideas on the objective test techniques that can be used in the exams.

The results of the last two sub-questions of the **RQ4** suggest that instructors' education degrees make no significant difference in the test techniques that they use in their classrooms. Moreover, whether the instructors have Master Degree or not does not determine the way they prefer the test techniques to be used in the exams. Thus, while organizing the members of the testing office, it does not make any difference whether the members have master degrees or not.

### **4.3 SUMMARY**

In this chapter, the findings of the statistical analysis of the study were presented. The research questions were given. Tables were formed including the findings of the study. Discussions about the findings were made under those related tables.

## **CHAPTER FIVE**

### **CONCLUSIONS, SUGGESTIONS AND IMPLICATIONS**

#### **5.0 INTRODUCTION**

In this final chapter, firstly a brief summary of the study is presented. Then, the conclusions which were drawn according to the results of the study are highlighted. Then, suggestions for the instructors, for the test constructors are made. Eventually, implications for further study are presented.

#### **5.1 SUMMARY OF THE STUDY**

In this study, it was aimed to find out the students' and the instructors' perceptions of the Compulsory English Language Course exams prepared by the testers of the testing office at COMU. The study also aimed to explore what other objective test techniques can also be used in these exams in the light of the students' and the instructors' ideas.

Quantitative research methodology was followed in this study. In order to achieve the aims above, the researcher formed some research questions after reviewing the related literature. To get the answers to those research questions, two questionnaires were prepared by the researcher in the light of the literature. Before piloting the questionnaires, the researcher consulted the ideas of three experts in 'English Language Teaching' and one expert in 'Measurement and Evaluation'. Having consulted the experts and piloted the questionnaires, the researcher omitted some parts of the questionnaires and added some extra items. With the help of the experts' ideas and the pilot study, the questionnaires got ready to be administered for the main study.

Main Study was conducted to 367 students and 33 instructors at COMU. Data needed were collected by using of the questionnaires. The researcher administered

the instructors' questionnaire to the instructors working at COMU and students' questionnaire to the students from eight different faculties and a college. All the students involving in the study were all in their second years at the university. Thus, they were all experienced about the testing system of the 'Compulsory English Language Course'.

The data obtained through the questionnaire were analyzed via Descriptive Statistics, One-way ANOVA, Independent Samples T-Test, Cronbach Alpha Reliability Test and Nonparametric Kruskal-Wallis Test by using SPSS (Statistical Package for Social Sciences) 13.0 for Windows. The results of the analysis were presented and discussed in Chapter Four named *Findings and Discussions*.

Some important conclusions were drawn and suggestions for better testing of students' language performance at university level were made taking the results of the study into consideration. Finally, implications for further study were introduced in the last part of the thesis.

## **5.2 CONCLUSIONS**

The answers to the four main research questions and the results of the study let the researcher draw the following conclusions:

While students want the instructors improve their vocabulary, pronunciation and speaking skills; instructors want to improve their students' grammar, vocabulary and reading skills respectively. There are significant differences among most of the language skills and areas that the instructors use and that the students expect to be used in their classes. As Dalyan (1990:113) claims, "language teaching does not mean to provide students with a mere knowledge of vocabulary and grammar. Students should also be furnished with all elements and skills of the language". The results verify Dalyan's statement. Students seem to need communicative lessons more than the lessons consisting of only grammar.

Most of the instructors (0.727) want the exams to be prepared by the testers of the testing office. However, more than the half of the students (0.657) prefers the instructor of the course to the testers of the testing office. Instructors' most common reason to prefer testing office is that they want all the instructors to follow the same curriculum. As for the students, they prefer the instructors of the course because they commonly think that the instructors should prepare their own exams by taking their students' different levels into consideration. Although most of the students want the exams to be prepared by the instructors of the course, they have some doubts about those exams' reliability.

Answers to the third research question show that both the instructors and the students have some doubts about the efficiency of the testing office's current practices. Students' ideas on the exams different characteristics differ in terms of their success. The more successful the students are, the more satisfied they are with the exams' different characteristics. As for the instructors; it has been found that the more experienced they are, the more efficient they find the exams prepared by the testing office.

Although multiple-choice test items are the fifth frequently used technique by the instructors, they are currently the only way to assess our learners' language performance. According to the results obtained from the answers to the fourth research question, instructors respectively prefer (1) multiple-choice questions, (2) matching, (3) ordering tasks, (4) completion, (5) true-false questions, (6) short-answer questions, (7) error correction and (8) word changing to be used in the SATs. Although their order of preferring objective test techniques slightly differs, the first eight test techniques that the students and instructors prefer to be used in the exams are completely same. The results are very similar to that of Dalyan's (1990) study. In his study, he also found that the most appropriate test techniques in the opinions of the instructors were respectively (1) multiple-choice questions, (2) matching, (3) true-false questions and (4) fill in the blanks (completion). This shows that these techniques are quite appropriate to be used at SATs. Finally, another finding of the



study is that instructors' having master degrees do not have any effect on their preferences of the objective test techniques.

Bearing these conclusions in mind, some valuable suggestions were made in the following parts.

### **5.3 SUGGESTIONS**

Taking the students' and the instructors' ideas on the Compulsory English Language Course exams into the consideration, following suggestions can be made for better testing of the students' language performance. Some of the suggestions were made by the instructors as they answered the open-ended questions given at the pilot study. As it was promised to be kept confidential, their names are not specified here. With the help of the suggestions for the instructors and for the test constructors, better testing practices are hoped to be achieved.

#### **5.3.1 SUGGESTIONS FOR THE INSTRUCTORS**

Following suggestions will help the instructors achieve better teaching and testing of English:

- 1.** Instructors should focus on the activities improving students' speaking, listening and writing skills. They must also do some pronunciation practices.
- 2.** Before the final and mid-term exams, an effective diagnostic test should be applied to the students as Heaton (1988) finds it really unfortunate not to have an effective diagnostic test which helps learners to see their weaknesses.
- 3.** In addition to the SATs of Compulsory English Language Course, instructors should also use teacher-made achievement tests as students have

different language levels. Burke's (2005:33) statement on this issue is well-founded. To him, "neither standardized tests alone, nor teacher[-made] assessments alone can provide a true picture of students' learning". Thus, instructors should also take the scores that students get from teacher-made achievement tests into consideration while assessing their learners' language performance.

### 5.3.2 SUGGESTIONS FOR THE TEST CONSTRUCTORS

The suggestions listed below will help the testers of the testing office prepare better and accurate tests:

1. According to the results of the study, both the students and the instructors prefer other test techniques to be used in SATs. Therefore, testers can include the eight objective test techniques, which were commonly preferred by the instructors and the students, into the SATs [(1) multiple-choice questions, (2) matching, (3) ordering tasks, (4) completion, (5) true-false questions, (6) short-answer questions, (7) error correction and (8) word changing]. As H. D. Brown (2001) believes, a test is a method which consists of different techniques, procedures and items. Similarly, in the opinion of Gordon (1998:11), "it is important to understand ... that no single assessment method can completely measure a student's range of skills and knowledge... Thus, it is necessary to use several types of assessment methods to help students learn..." Thus, Centrally Administered Achievement test of Compulsory English Language Course cannot be carried out using only multiple-choice test items. These tests should include other test techniques as well.

2. As suggested by some of the instructors; although it is not much objective to be scored, *translation* should also be included in SATs.

3. As the instructors and the students do not agree that the questions used in the exams are authentic, test constructors should ask the instructors to prepare more authentic and meaningful questions to be used in the SATs of COMU.
4. The instructors devote most of their time to teaching. They do not seem to have enough time to spend on testing. Because of this, the questions that they prepare and send to the testers of the testing office usually have some problems. Nevertheless, administering a test is a specific skill which needs to be considered as important. Hence, having a testing office with testers specialized just at testing will be a good idea for preparing better and more efficient tests. Although this testing office seems to be independent; with good coordination among the instructors and the testers, it will be beneficial both for the students and for the instructors; thus, for language learning.
5. As it was also suggested by Şahinel (1997), the testing office and the curriculum development unit should work together for better testing.
6. Finally, a computer-based measuring and assessing system should be founded for the testing office to have a profitable follow-up procedure as suggested by one of the instructors of COMU.

#### **5.4 IMPLICATIONS FOR FURTHER STUDY**

As this study is limited to a particular setting and to particular participants, further study is needed. The implications listed below will provide useful facts for the further study:

1. This research only covered the students' and the instructors' ideas on SATs at COMU. A similar study can be conducted at other universities having SATs for their Compulsory English Language Courses.

2. The ways of integrating all skills and areas both to the courses and to the exams can be explored. Thus, scales for the use of both subjective and objective tests at SATs can be developed.

3. Besides the fourteen test techniques listed in the questionnaires, further test techniques – both objective and subjective ones – can be incorporated in the analysis of achievement tests so that research can be enlarged.

4. In addition to using only questionnaires, new techniques can also be used to get more information about the current testing practices.

5. In the light of the data collected and analyzed in this study, a model test can be developed and applied to a group of students randomly chosen. Then, the differences between the effects and results of the new model test and the current testing system can be questioned.

6. As the results of the study show that the more successful the students are, the more satisfied they are with the exams' different characteristics, it will be good idea to get the number of the successful, unsuccessful, and very successful students before the new study is conducted and form the samples of the further studies accordingly.

7. Finally, it has been found that having a master degree did not have any effect on the instructors' choices of the particular test techniques. What might affect the instructors' choices of the test techniques can be a good question for further study.

## REFERENCES

AHMANN, J.S. and M. GLOCK

1971 *Measuring and Evaluating Educational Achievement*.  
Boston, MA: Allyn and Bacon.

AKSAN, M. Metin

2001 "Instructors Perceptions of the Content Validity of the English Language Exams at Niğde University".  
Unpublished MA thesis. ANKARA: Bilkent University.

ALDERSON, J. Charles

2000 *Assessing Reading*.  
CAMBRIDGE: Cambridge University Press.

ALDERSON, J. C., C. CLAPHAM and D. WALL

1995 *Language Test Construction and Evaluation*.  
CAMBRIDGE: Cambridge University Press.

ANASTASI, Anne

1982 *Psychological Testing*.  
LONDON: Macmillan

BACHMAN, LYLE F.

1990 *Fundamental Considerations in Language Testing*.  
OXFORD: Oxford University Press.

BACHMAN, L. F. and A. S. PALMER

1996 *Language Testing in Practice*.  
OXFORD: Oxford University Press.

BAKER, David

1989 *Language Testing: A Critical Survey and Practical Guide*.  
London : Edward Arnold.

BALCI, Ali

2005 *Sosyal Bilimlerde Araştırma: Yöntem Teknik ve İlkeler* (5. Baskı).  
Ankara: Pegem A Yayıncılık.

BELL, Judith

1993 *Doing Your Research Project: A Guide for First-time Researchers in Education and Social Science*.  
BUCKINGHAM: Open University Press.

BOBDA, A. S.

1993 Testing Pronunciation.  
*English Teaching Forum Online*. 31(3), 18-22. Available at:  
<http://exchanges.state.gov/forum/vols/vol31/no3/p18.htm>

BOSTAN, Sibel

2005 "EFL Teachers' Effectiveness in Testing and Evaluating Student Performance in English as a Foreign Language"  
Unpublished MA thesis. ÇANAKKALE: Çanakkale Onsekiz Mart University.

BRACEY, G. W.

1998 *Put to the Test: An Educators' and Consumers' Guide to Standardized Testing*.  
Bloomington, IN: Center for Professional Development and Services,  
Phi Delta Kappa International.

BROWN, H. Douglas

2001 *Teaching by Principles: An Interactive Approach to Language Pedagogy.*

New York, NY: Addison Wesley Longman.

BROWN, James Dean

1996 *Testing in Language Programs.*

New Jersey: Prentice Hall Regents.

BURKE, Kay

2005 *How to Assess Authentic Learning.*

Thousand Oaks, California: Sage Publications.

BÜYÜKÖZTÜRK, Şener

2006 *Sosyal Bilimler için Veri Analizi El Kitabı (6. Baskı).*

Ankara: Pegem A Yayıncılık.

CAMBRIDGE LEARNER'S DICTIONARY

2004 Cambridge: Cambridge University Press.

CELCE-MURCIA, Marianne

2001 *Teaching English as a Second or Foreign Language (3rd ed.).*

Boston MA: Heinle & Heinle.

COHEN, A. D.

1994 *Assessing Language Ability in the Classroom.*

BOSTON: Heinle & Heinle.

COHEN, A. D.

2001 *Second Language Assessment.*

In Celce-Murcia, M. *Teaching English as a second or foreign language. (3<sup>rd</sup> ed.).* Boston MA: Heinle & Heinle.

DALYAN, Ateş

1990 “The Comments of EFL Teachers on Preparing Achievement Tests and an Analysis of a Sample Test (A Study at the Engineering Faculty of Anadolu University).”

Unpublished MA thesis. ESKİŞEHİR: Anadolu University.

DAVIES, Alan

1975 Two Tests of Speeded Reading.

In R. L. Jones and B. Spolsky (eds.), *Testing Language Proficiency*.

Washington, DC: Center for Applied Linguistics.

DAVIES, Alan

1990 *Principles of Language Testing*.

OXFORD: Basil Blackwell.

DAVIES, A., A.BROWN, C. ELDER, K. HILL, T. LUMLEY and T. MCNAMARA

1999 *Studies in Language Testing: Dictionary of Language Testing*.

Cambridge: Cambridge University Press.

DEMİREL, Özcan

2002 *Öğretme sanatı*.

Ankara: Pegem A Yayıncılık.

DICKINS, R. and K. GERMAINE

1995 Evaluation.

*ELT Journal*. 49(2), 194-195.

DORNYEI, Zoltan

2003 *Questionnaires in Second Language Research: Construction, Administration, and Processing*.

New Jersey: Lawrence Erlbaum Associates, Inc.



FROST, Richard

- 2005 *Test Question Types*. Available at:  
<http://www.teaching-english.org.uk>. Turkey: British Council

GENESEEE, F. and J. A. UPSHUR

- 1996 *Classroom-based Evaluation in Second Language Education*.  
CAMBRIDGE: Cambridge University Press.

GORDON, H. R. D.

- 1998 "Vocational education teachers' perception of their use of  
assessment methods".  
*Journal of Vocational and Technical Education*. 15(1): 1 - 14

GREEN, J. A.

- 1970 *Introduction to Measurement and Evaluation*.  
New York, NY: Dodd, Mead & Company.

GRONLUND, N.E.

- 1971 *Measurement and Evaluation in Teaching*.  
London: Macmillan.

HARMER, Jeremy

- 2001 *The Practice of English Language Teaching* (Third Ed.).  
Malaysia: Pearson Education Limited.

HEATON, John Bright

- 1988 *Writing English Language Tests*.  
LONDON: Longman.

HUGHES, Arthur

- 1989 *Testing for Language Teachers*.  
CAMBRIDGE: Cambridge University Press.

KITAO, S. Kathleen and K. KITAO

1996 Testing Grammar.

*The Internet TESL Journal*, Vol. 2, No. 6 Available at:

<http://iteslj.org/Articles/Kitao-TestingGrammar.html>

KÖKSAL, Dinçay

2004 “Assessing Teachers’ Testing Skills in ELT and Enhancing Their Professional Development Through Distance Learning on the Net.”

*Turkish Online Journal of Distance Education TOJDE*, 5 (1): 1-6.

KÖKSAL, D. and S. BECEREN

2003 “Do we need multiple choice tests in language teaching and testing?”

Paper presented at the Sofia Conference. Bulgaria: Sofia.

KUNTASAL, İlknur

2001 “Perceptions of Teachers and Testers of Achievement Tests Prepared by Testers in the Department of Basic English at the Middle East Technical University.”

Unpublished MA thesis. ANKARA: Bilkent University.

MADSEN, S. Harold

1983 *Techniques in Testing*.

Oxford: Oxford University Press.

MCNAMARA, Tim

2000 *Language Testing*.

Oxford: Oxford University Press.

MURPHY, Dermot

2000 Key Concepts in ELT.

*ELT Journal*, 54(2): 210-211. OXFORD: Oxford University Press.

ÖSKEN, Hilal

- 1999 "An Assessment of the Validity of the Midterm and the End of Course Assessment Tests Administered at Hacettepe University, Department of Basic English."  
Unpublished MA thesis. ANKARA: Bilkent University.

PARSONS, J. and T. FENWICK

- 1999 "Using Objective Tests to Evaluate".  
*Education Resources information Centre*. (ED431013). Available at:  
[http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/17/97/f5.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/17/97/f5.pdf).

POPHAM, W. J.

- 2003 *Test better, teach better: The instructional role of assessment*.  
USA: Association for Supervision and Curriculum Development.

RAATZ, U.

- 1985 Tests of reduced redundancy: The C-test, a practical example.  
*Fremdsprachen und Hochschule*. Bochum: AKS Rundbrief. 13/14:  
14-19.

RUDMAN, H. C.

- 1989 "Integrating testing with teaching".  
*Eric Clearinghouse on Tests Measurement and Evaluation*  
(ED315432). (On-line). Available:  
[www.ed.gov/databases/ERIC\\_Digests/ed315432.html](http://www.ed.gov/databases/ERIC_Digests/ed315432.html)

SALKIND, N. J.

- 2006 *Tests and Measurement for People Who (Think They) Hate Tests and Measurement*.  
Thousand Oaks, California: Sage Publications.

SERPIL, Harun

2000 “An Assessment of the Content Validity of the Midterm Achievement Tests Administered at Anadolu University Foreign Languages Department.”

Unpublished MA thesis. ANKARA: Bilkent University.

SMITH, F.M. and S. ADAMS

1972 *Educational Measurement for the Classroom Teacher.*

New York, NY: Harper & Row.

ŞAHİNEL, Semih

1997 “Ankara Üniversitesi Hazırlık Sınıflarında Uygulanmakta Olan İngilizce Sınavlarına İlişkin Okutmanlarının Görüşleri.”

Unpublished MA thesis. ANKARA: Hacettepe University.

TATAROĞLU, N.

1995 *100 Reading Passages with Test Questions.*

Ankara: Saypa Yayınları.

THORNBURY, S.

1999 *How to Teach Grammar.*

Essex: Pearson Education.

URDAN, T. and S. PARIS

1994 “Teachers' perceptions of standardized achievement tests”.

*Educational Policy.* 8(2): 137-156.

VERGILI, Atanur

1984 “Teachers’ Attitudes toward Testing at METU Gaziantep Preparatory School.”

Unpublished MA thesis. ANKARA: Middle East Technical University.

WEIR, Cyril J.

1990 *Communicative Language Testing*.  
New York: Prentice Hall.

WIGGINS, G.

1989 "Teaching to the (authentic) test".  
*Educational Leadership*. 49(8). 26-33.

## **APPENDICES**

**APPENDIX A:** Open-ended Questions to Construct the Third Part of the Students' Questionnaire

**APPENDIX B:** Open-ended Questions to Construct the Third Part of the Instructors' Questionnaire

**APPENDIX C:** Piloting the Questionnaire

**APPENDIX D:** Students' Questionnaire for the Main Study

**APPENDIX E:** Instructors' Questionnaire for the Main Study

## APPENDIX A

Sevgili Öğrenciler,

Bu çalışma tamamıyla bilimsel amaçlı olup, vereceğiniz cevaplar gizli tutulacaktır. Bu yüzden isminizi belirtmenize gerek yoktur. Ankete vereceğiniz samimi cevaplar çalışmama büyük katkı sağlayacaktır. Yardımınız için şimdiden teşekkür ederim.

Kürşat CESUR

Çanakkale Onsekiz Mart Üniversitesinde birinci sınıfta almış olduğunuz Zorunlu İngilizce Dersi vize ve finalinde merkezi olarak hazırlanan ve tüm Çanakkale’de ortak olarak uygulanan bir sınav kullanılıyor. Sizce bu sınavlar şu anda olduğu gibi merkezi olarak mı yürütülsün, yoksa dersin sorumlu öğretim elemanı tarafından mı hazırlansın? Nedenlerini Belirtiniz. İkisinin de avantaj ve dezavantajları olduğunu düşünüyorsanız nedenlerinizi iki bölüme de yazabilirsiniz.

1. Çanakkale Onsekiz Mart Üniversitesinde birinci sınıfta almış olduğum Zorunlu İngilizce Dersinin Sınavı şu anda olduğu gibi merkezi olarak yürütülsün. Çünkü:

- 
- 
- 
- 
- 

2. Çanakkale Onsekiz Mart Üniversitesinde birinci sınıfta almış olduğum Zorunlu İngilizce Dersinin Sınavı dersin sorumlu öğretim elemanı tarafından hazırlansın. Çünkü:

- 
- 
- 
- 
-

## APPENDIX B

Değerli Meslektaşım,

Bu çalışma tamamıyla bilimsel amaçlı olup, vereceğiniz cevaplar gizli tutulacaktır. Bu yüzden isminizi belirtmenize gerek yoktur. Ankete vereceğiniz samimi cevaplar çalışmama büyük katkı sağlayacaktır. Yardımınız için şimdiden teşekkür ederim.

Kürşat CESUR

Bildiğiniz gibi, Çanakkale Onsekiz Mart Üniversitesinde; birinci sınıfta vermiş olduğunuz Zorunlu İngilizce Dersi sınavları merkezi olarak hazırlanıp, tüm Çanakkale’de ortak olarak yürütülüyor. Sizce bu sınavlar şu anda olduğu gibi merkezi olarak mı yürütülsün, yoksa dersin sorumlu öğretim elemanı tarafından mı hazırlansın? Nedenlerini Belirtiniz. İkisinin de avantaj ve dezavantajları olduğunu düşünüyorsanız nedenlerinizi iki bölüme de yazabilirsiniz.

1. Zorunlu İngilizce Dersinin Sınavı merkezi olarak yürütülsün. Çünkü:

- 
- 
- 
- 
- 
- 
- 

2. Zorunlu İngilizce Dersinin Sınavı dersin sorumlu öğretim elemanı tarafından hazırlanıp, yürütülsün. Çünkü:

- 
- 
- 
- 
- 
- 
-



## APPENDIX C

### PILOTING THE QUESTIONNAIRE

1. How long did it take you to complete this questionnaire?

---

2. Were the instructions clear?

---

3. Were any of the questions / sentences unclear or ambiguous? If so, can you write which and why?

---

---

4. Are there any words or phrases that you want to be explained? If yes, can you please write them?

---

---

---

5. Did you object to answering any of the questions?

---

---

6. In your opinion, has any major topic been omitted? If yes, can you please write which topics should also be included in this questionnaire?

---

---

7. Was the layout of the questionnaire clear/attractive?

---

8. Any comments?

---

---

---

---

---

---

(Adapted from Bell, 1993)

## APPENDIX D

### ÜNİVERSİTE SEVİYESİNDE DİL BECERİLERİNİ DEĞERLENDİRMEDE KULLANILAN OBJEKTİF TEST TEKNİKLERİNE İLİŞKİN ÖĞRENCİ VE OKUTMANLARIN ALGILAMALARI

#### Öğrenci Anketi

Değerli Öğrenciler,

Bu anketin amacı öğrencilerin “Zorunlu İngilizce Dersi Merkezi Sınavlarına” ilişkin algılamaları hakkında fikir sahibi olmaktır. Ayrıca bu çalışma öğrencilerin görüşlerini dikkate alarak, çoktan seçmeli soru tipinden başka ne tür objektif soru tiplerinin kullanılabilceğini tespit etmeyi amaçlamaktadır.

Ankete vereceğiniz cevaplar Zorunlu İngilizce Derslerindeki Ölçme-Değerlendirme sürecinin etkililiği hakkında bilgi verecek ve çalışmaya büyük katkı sağlayacaktır. Bu çalışma tümüyle bilimsel amaçlı olup vereceğiniz cevaplar araştırma dışında hiçbir amaçla kullanılmayacak ve gizliliğe tam bir özen gösterilecektir. Bu nedenle isim belirtmenize gerek duyulmamaktadır.

Anket sorularını içtenlikle cevaplayacağınız için şimdiden teşekkür ederim.

İngilizce Okutmanı  
Kürşat CESUR

<b>BÖLÜM -1-: Sizin için uygun olan seçeneğe çarpı (X) işareti koyunuz.</b>					
<b>Cinsiyet:</b>	a. Bay ( )		b. Bayan ( )		
<b>Birinci sınıftaki “Zorunlu İngilizce Dersi” Bahar Dönemi harf notunuz:</b>					
a. AA ( )	b. BA ( )	c. BB ( )	d. CB ( )	e. CC ( )	
f. DC ( )	g. DD ( )	h. FD ( )	i. FF ( )	j. DS ( )	
<b>Eğitim gördüğünüz Fakülte ya da Yüksek Okul:</b>					
a. Eğitim F. ( )	b. Fen Edebiyat F. ( )		c. Güzel Sanatlar F. ( )		
d. Mimarlık Mühendislik F. ( )	e. İlahiyat F. ( )		f. Biga İktisadi ve İdari Bilimler F. ( )		
g. Su Ürünleri F. ( )	h. Ziraat F. ( )		i. Sağlık YO ( )		

<b>BÖLÜM -2-: Sizin için uygun olan kutuya çarpı (X) işareti koyunuz.</b>					
• Aşağıda verilen farklı dil becerilerini ve alanlarını geliştirecek aktivitelerin Zorunlu İngilizce derslerinde ne kadar sıklıkta kullanılmasını <b>isterdiniz?</b>					
<b>Daima = “5” / Genellikle = “4” / Bazen = “3” / Nadiren = “2” / Hiç = “1”</b>					
<b>Dil Becerileri ve Alanları</b>	<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>1</b>
Dinleme (Listening) Öğretimi					
Konuşma (Speaking) Öğretimi					
Okuma (Reading) Öğretimi					
Yazma (Writing) Öğretimi					
Dil Bilgisi (Grammar) Öğretimi					
Kelime (Vocabulary) Öğretimi					
Telaffuz (Pronunciation) Öğretimi					

**BÖLÜM -3-: Sizin için uygun olan seçeneğe ya da kutuya çarpı (X) işareti koyunuz.**

Zorunlu İngilizce Dersi sınavları merkezi olarak aynı anda yapılmakta ve ölçme değerlendirme birimi tarafından organize edilmektedir. Sizce bu sınav şu anda olduğu gibi bir birim tarafından mı yoksa dersin öğretim elemanı tarafından mı hazırlanıp, yürütülsün?

1. Ölçme Değerlendirme Birimi organize etsin: ( ) 2. Dersten sorumlu Öğretim Elemanı organize etsin: ( )

• Yukarıda birinci maddeyi işaretlediyseniz yalnızca “A” bölümünü; ikinci maddeyi işaretlediyseniz yalnızca “B” bölümünü cevaplayınız.

**Kesinlikle katılıyorum = 5, Katılıyorum = 4, Kararsızım = 3, Katılmıyorum = 2, Kesinlikle katılmıyorum = 1**

A) Sınavları Ölçme Değerlendirme Birimi organize etsin çünkü...	5	4	3	2	1
1. Merkezi sınav farklı bölümlerdeki öğrenciler arasındaki seviye farkını daha iyi belirler.					
2. Sınavlarda çoktan seçmeli (dört seçenekli) soru tipi kullanılmaktadır.					
3. Merkezi sınav tüm öğrencilerin bilgisini adil olarak ölçer.					
4. Merkezi sınav daha güvenilirdir.					
5. Öğretim Elemanının sınıfa karşı olası olumsuz tutumunun sınav sorularına yansımaları engeller.					
6. Diğer: Varsa belirtiniz...					
B) Sınavları Dersten sorumlu Öğretim Elemanı organize etsin çünkü...	5	4	3	2	1
1. Öğrenciler arasındaki seviye farkını bildiği için, öğretim elemanı bu farkı dikkate alarak sorular hazırlar.					
2. Sınavda çoktan seçmeli test tekniğinin yanında derste işlenen diğer soru tipleri de kullanılabilir.					
3. Öğrencilerin seviyeleri farklı olduğu için, öğretim elemanın kendi sorularını hazırlaması öğrenciler açısından daha adildir.					
4. Öğretim elemanının hazırladığı sınav daha güvenilirdir.					
5. Öğretim Elemanının sınıfa karşı olası olumlu tutumu değerlendirme sürecini olumlu etkiler.					
6. Diğer: Varsa belirtiniz...					

**Kesinlikle katılıyorum = 5, Katılıyorum = 4, Kararsızım = 3, Katılmıyorum = 2, Kesinlikle katılmıyorum = 1**

BÖLÜM -4-: Ölçme Değerlendirme Biriminin Hazırladığı Merkezi Sınavlar Hakkında Ne Düşünüyorsunuz?	5	4	3	2	1
1. Sınavlarda kullanılan sorular dersin amaçlarıyla uyum içindedir.					
2. Sınavların içeriği, sınıfta öğrendiğim ders içeriğiyle uyum içindedir.					
3. “Çoktan Seçmeli” soru tipi sınıfta kullanılan aktivite türlerine uymaktadır.					
4. Sınavlarda sorulan sorular derste öğrendiğim konuyu iyi bir şekilde test eder.					
5. Çoktan seçmeli test tekniği dil performansımı ölçmede <b>başarılıdır</b> .					
6. Çoktan seçmeli test tekniği dil performansımı ölçmede <b>yeterlidir</b> .					
7. Sınavda özgün sorular kullanılmaktadır.					
8. Bu sınavlarda dil öğretildiği gibi değerlendirilmektedir.					
9. Sorular anlayabilmemiz için yeterince açıktır.					
10. Sınavların tasarımı öğrenciler için uygundur.					
11. Sınav organizasyonu yeterlidir.					
12. Sınavlarda birçok test türü olacağına, sadece çoktan seçmeli soru tipi olmalıdır.					

**BÖLÜM -5-**  
**OBJEKTİF TEST TEKNİKLERİ**

Aşağıdaki objektif test tekniklerinden bilmediğiniz varsa, her bir test tekniğinin altında açıklaması bulunmaktadır. Lütfen bu açıklamaları dikkatle okuyarak sizin için uygun olan kutulara çarpı (X) işareti koyunuz (A ve B sütunu olmak üzere, her bir teknik için iki tane çarpı (X) işareti olmalıdır).

**Sütun (A):** Öğretim elemanları bu teknikleri ders içi aktivitelerde ne kadar sıklıkta kullanıyor?

**Sütun (B):** Bu tekniklerden hangileri Zorunlu İngilizce Dersi merkezi sınavlarında kullanılabilir?

OBJEKTİF TEST TEKNİKLERİ	(A) SINIFTA KULLANILILIR					(B) SINAVDA KULLANILABİLİR				
	DAİMA	GENELLİKLE	BAZEN	NADİREN	HİÇ	KESİNLİKLE KATILYORUM	KATILYORUM	KARARSIZIM	KATILMIYORUM	KESİNLİKLE KATILMIYORUM
<b>1. Çoktan Seçmeli Sorular:</b>										
Bir soru kökü ve çoktan seçmeli maddelerden oluşan soruya denir. Verilen dört ya da beş şıktan yalnızca biri doğrudur.										
<b>2. Kısa Cevaplı Sorular:</b>										
<b>3. Doğru-Yanlış Soruları:</b>										
<b>4. Karşılaştırma:</b>										
Öğrencilerin iki ayrı sütunda verilen sorular ile cevapları, kelimelerle anlamları, resimlerle kelimeleri karşılaştıracakları soru tipidir.										
<b>5. Boşluk Doldurma:</b>										
Tamamlanmamış bir cümleye cevap olarak bir kelime ya da kelime grubu gerektiren soru tipidir.										
<b>6. Cloze Test:</b>										
Paragraftaki her beşinci, yedinci, dokuzuncu, vb. kelimelerin yerine boşluk bırakılır. Öğrenciler uygun bir şekilde boşlukları doldururlar.										
<b>7. C-Test:</b>										
Cloze teste benzer. Ancak cloze teste olduğu gibi bir kelimenin tamamını silmek yerine, her ikinci kelimenin ikinci yarısı silinmektedir.										
<b>8. Cloze Elide Test:</b>										
Parçadan kelimeleri silmek yerine parçaya yeni kelimeler eklenir. Öğrencilerden bu eklemelerin nerede yapıldığını bulmaları istenir.										
<b>9. Sıraya Koyma:</b>										
Karışık kelimeler, cümleler ya da paragraflar verilir. Öğrencilerden bunları doğru sıraya koymaları istenir.										
<b>10. Hata Düzeltme:</b>										
İçinde öğrencilerin bulmaları ve düzeltmeleri gereken hataların olduğu cümle veya okuma parçalarından oluşur.										
<b>11. Dönüştürme:</b>										
Bir cümle ve ikinci cümlenin ilk birkaç kelimesi verilip, öğrencilerden ilk cümlenin anlamı değişmeden ikincisini oluşturmaları istenir.										
<b>12. Birleştirme:</b>										
İki cümle ve parantez içinde bir kelime verilir. Öğrencilerden parantez içinde verilen kelimeyi kullanarak cümleleri birleştirmesi istenir.										
<b>13. Ekleme:</b>										
Öğrencilerden büyük harfle yazılan kelimeleri cümle içinde en uygun yere eklemeleri istenir. <b>Örn:</b> ALWAYS She helps her mother.										
<b>14. Kelime Değiştirme:</b>										
Öğrencilere tamamlanmamış bir cümle ve parantez içinde yapısını değiştirerek boşluğu doldurmaları gereken kelime verilir. Öğrenciler kelimeyi uygun yapıda boşluğa doldururlar. <b>Örn:</b> I have never _____ Chinese food. (eat)										
<b>15. Ekleme istediğiniz başka soru tipi varsa, lütfen belirtiniz.....</b>										
.....										

## APPENDIX E

### STUDENTS' AND INSTRUCTORS' PERCEPTIONS OF THE OBJECTIVE TEST TECHNIQUES USED TO ASSESS LANGUAGE SKILLS AT UNIVERSITY LEVEL

#### Instructors' Questionnaire

Dear Colleague,

The main purpose of this questionnaire is to have an idea on the instructors' perceptions the Compulsory English Language Course exams. Moreover, this study aims to determine what other objective test techniques can also be used in these exams in addition to the multiple-choice test technique while taking all the instructors' opinions into consideration.

I regard your answers as a valuable contribution to my study. Your answers will help me a great deal with my research. They will provide important information about the effectiveness of the testing practices at Compulsory English Language Courses. The answers to this questionnaire will be kept confidential. You do not have to write your name and no one will know your specific answers to this questionnaire.

Thank you for your kind co-operation in completing this questionnaire.

Kürşat CESUR

<b>PART -1-: Please put a cross (X) into the brackets which is appropriate for you. If there is another choice, please specify it into the "other" section.</b>					
<b>Sex:</b>	a. Male ( )	b. Female ( )			
<b>Teaching Experience:</b>	a. 0-3 years ( )	b. 4-6 years ( )	c. 7-9 years ( )	d. 10-12 years and more ( )	
<b>Have you ever taught at a Faculty or College?</b>	a. YES ( )		b. NO ( )		
<b>Graduation:</b>	a. BA Degree ( )	b. MA Degree ( )			
<b>Department:</b>					
a. English Language Teaching: BA ( ) / MA ( )    b. English Language and Literature: BA ( ) / MA ( )					
c. English Linguistics: BA ( ) / MA ( )    d. Translation and Interpretation: BA ( ) / MA ( )					
e. American Culture and Literature: BA ( ) / MA ( )    f. Other _____ (MA)					
<b>Have you attended to the things below on "Testing and Evaluation"? (More than one option is possible)</b>					
a. BA degree course ( )		b. MA degree course ( )		c. Seminar ( )	
d. Conference ( )		e. Symposium ( )		f. Others ( ) _____	

<b>PART -2-: Please put a cross (X) into the box which is appropriate for you.</b>					
Different language skills and areas to be taught are listed below. How often do you use classroom activities which help your students' different language skills and areas develop?					
Language Skills and Areas	<b>Always (5)</b>	<b>Usually (4)</b>	<b>Sometimes (3)</b>	<b>Rarely (2)</b>	<b>Never (1)</b>
Teaching Listening					
Teaching Speaking					
Teaching Reading					
Teaching Writing					
Teaching Grammar					
Teaching Vocabulary					
Teaching Pronunciation					



**PART -5-  
OBJECTIVE TEST TECHNIQUES**

Some objective test techniques are listed below. If you are not familiar with some of these techniques, you may read the descriptions of them in the last part of the questionnaire (GLOSSARY).

Please put a cross (X) into the box which is appropriate for you.

You should put two crosses for each technique as there are two different columns to be answered.

**Column (A):** How often do you use these techniques in your classroom activities?

**Column (B):** Which of these techniques can be used in centrally administered achievement tests?

OBJECTIVE TEST TECHNIQUES WHICH	(A) ARE USED IN MY CLASS					(B) CAN BE USED IN OUR EXAMS				
	ALWAYS	USUALLY	SOMETIMES	RARELY	NEVER	STRONGLY AGREE	AGREE	UNDECIDED	DISAGREE	STRONGLY DISAGREE
1. Multiple-Choice Questions										
2. Short-Answer Questions										
3. True-False Questions										
4. Matching										
5. Completion										
6. Cloze Test										
7. C-Test										
8. Cloze Elide Test										
9. Ordering Tasks (Rearrangement)										
10. Error Correction										
11. Transformation										
12. Combination										
13. Addition										
14. Word Changing										

Please specify below if there are any other techniques you would like to add.

15.

16.

**PART -6-**  
**GLOSSARY**

**Addition:** Tester can ask the students to insert the words in capitals in the most appropriate place in each sentence as in the following example:

e.g. ALWAYS She helps her mother.

**Centrally Administered Achievement Test:** Any assessment device that is administered and scored in a standard, predetermined manner.

**Cloze Test:** A true cloze test is a text in which every *-n*th word has been deleted. There is an example of a cloze test in which every seventh word was deleted.

e.g. Preparations are underway for the Pan-World Games in Lomoka next year. Many new hotels have (1) \_\_\_\_\_ been built and tourists are already (2) \_\_\_\_\_ reservations. But the main stadium hasn't (3) \_\_\_\_\_ started yet. They are still deciding (4) \_\_\_\_\_ to put it. The Athletes' Village (5) \_\_\_\_\_ still being built, and the swimming (6) \_\_\_\_\_ isn't completed yet.

**Cloze Elide Test:** In this test technique, the tester inserts words instead of deleting them. Students have to show where these insertions have been made.

e.g. Aşure is with the name of a pudding which is been cooked with extremely diverse and seemingly disharmonious ingredients. It is a pudding through that has become a part of many the Muslim tradition. It is believed to have originated from when Noah's Ark came to rest on a mountain of the Turkey after the great floods...

**C-Test:** This test technique is similar to the cloze tests. However, instead of whole words as in cloze tests, second half of every second word is deleted in this technique.

e.g. Pollution is one of the big problems in the world today. Towns a\_\_ cities a\_\_ growing, indu\_\_ is gro\_\_ and t\_\_ population o\_\_ the wo\_\_ is gro\_\_. Almost every\_\_ causes poll\_\_ in so\_\_ way o\_\_ another. T\_\_ air i\_\_ filled wi\_\_ fumes fr\_\_ factories a\_\_ vehicles, a\_\_ there i\_\_ noise fr\_\_ airplanes a\_\_ machines. Riv\_\_, lakes, a\_\_ seas a\_\_ polluted b\_\_ factories and by sewage from our homes.

**Combination:** Tester can ask the students to combine the sentences using the words in brackets as in the following example.

e.g. You finish your homework. Then, check it carefully. (After)

**Error Correction:** These tests usually consist of sentences or passages in which there are some errors that the students have to identify and correct.

**Objective Test:** A test can be called as objective or subjective by determining the way how the teacher scores the students' performances.

**Ordering Tasks:** Students are given scrambled set of words, sentences, paragraphs or texts. They have to put them into their correct order.

**Transformation:** The tester gives a sentence and the first few words of another sentence. Then, he/she asks students to change the original sentence without changing the meaning as it is seen in the example below:

e.g. In sunny weather, I often go for a walk.  
When the weather \_\_\_\_\_.

**Word Changing:** The students are given an incomplete sentence and a word which they need to complete the sentence by changing the form of this word and by inserting it into the sentence in its correct form as in the examples below.

e.g. I have never \_\_\_\_\_ Chinese food. (eat)  
The flowers should be \_\_\_\_\_ everyday. (water)