

AKDENIZ UNIVERSITY
INSTITUTE OF EDUCATIONAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGE EDUCATION
PROGRAM OF ENGLISH LANGUAGE TEACHING

**EXAMINING THE RELATIONSHIP BETWEEN ENGLISH ADMISSION
TEST (LYS-5) AND FOUR LANGUAGE SKILLS OF
UNDERGRADUATE STUDENTS**

MASTER'S THESIS

Burak ASMA

Antalya

July, 2018

AKDENİZ UNIVERSITY
INSTITUTE OF EDUCATIONAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGE EDUCATION
PROGRAM OF ENGLISH LANGUAGE TEACHING

**EXAMINING THE RELATIONSHIP BETWEEN ENGLISH ADMISSION
TEST (LYS-5) AND FOUR LANGUAGE SKILLS OF
UNDERGRADUATE STUDENTS**

MASTER'S THESIS

Burak ASMA

Supervisor:

Dr. Fatma Özlem SAKA

Antalya

July, 2018

ATTESTATION

I honorably certify that the current study presented as my MA dissertation reports original work by me and was written within the scope of the ethics of research. I verify that the references contain the sources I utilized for my research through citation. I notify that I will bear all the ethical and legal consequences in condition that it is refuted.

DOĞRULUK BEYANI

Yüksek Lisans tezi olarak sunduğum bu çalışmayı, bilimsel ahlak ve geleneklere aykırı düşecek bir yol ve yardıma başvurmaksızın yazdığımı, yararlandığım eserlerin kaynakçalarda gösterilenlerden oluştuğunu ve bu eserleri her kullanımda alıntı yaparak yararlandığımı belirtir; bunu onurumla doğrularım. Enstitü tarafından belli bir zamana bağlı olmaksızın, tezimle ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara katlanacağımı bildiririm.

27/07/2018

Burak ASMA

T.C.
AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Burak ASMA 'nın bu çalışması 27.07.2018 tarihinde jürimiz tarafından **Yabancı Diller Eğitimi** Anabilim Dalı **İngiliz Dili Eğitimi** Tezli Yüksek Lisans Programında **Yüksek Lisans Tezi** olarak **oy birliği/oy çokluğu** ile kabul edilmiştir

İMZA

Başkan : (Doç. Dr.) **Turan PAKER**
(Pamukkale Üniversitesi, Eğitim Fakültesi,
Yabancı Diller Eğitimi Bölümü)



Üye : (Doç. Dr.) **Binnur GENÇ İLTER**
(Akdeniz Üniversitesi, Eğitim Fakültesi,
Yabancı Diller Eğitimi Bölümü)



Üye (Danışman) : (Dr.Öğr.Ü.) **Fatma Özlem SAKA**
(Akdeniz Üniversitesi, Eğitim Fakültesi,
Yabancı Diller Eğitimi Bölümü)



YÜKSEK LİSANS TEZİNİN ADI:

Examining the Relationship between English Admission Test (LYS-5) and Four Language Skills of Undergraduate Students

İngilizce Lisans Yerleştirme Sınavı (LYS-5) ile Lisans Öğrencilerinin Dört Dil Becerisi Arasındaki İlişkinin İncelenmesi

ONAY: Bu tez, Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulunun tarihli ve sayılı kararıyla kabul edilmiştir.

(Doç. Dr. Ramazan KARATAŞ)

Enstitü Müdürü

ACKNOWLEDGEMENTS

This thesis is not simply the outcome of my individual efforts but the cooperation of many people.

In this way, I would like to extend my heartfelt thanks;

- to the participants of the research for their sincere assistance and cooperation,
- to my supervisor, Dr. Fatma Özlem Saka, for her never-ending support and guidance throughout the thesis,
- to Assoc. Prof. Dr. Turan Paker for his pinpoint feedbacks and suggestions for my thesis.
- to Assoc. Prof. Dr. Binnur Genç İter for enlightening my path and giving a novel touch to my educational and academic life,
- to my lecturers, Dr. Mustafa Caner and Dr. Simla Course, for their support and the sources they shared,
- to Dr. Gizem Uyumaz and Merve Ayvallı for their priceless contributions,
- to my colleague, Orhan Ünal, for sharing his comments and supporting me during the process,
- to my parents for their prayers and eternal support every moment of my life,
- to my beloved wife for her non-stop encouragement and invaluable support throughout the process.

Burak ASMA

ABSTRACT

EXAMINING THE RELATIONSHIP BETWEEN ENGLISH ADMISSION TEST (LYS-5) AND FOUR LANGUAGE SKILLS OF UNDERGRADUATE STUDENTS

Asma, Burak

Master of Arts, Department of Foreign Language Education

Supervisor: Dr. Fatma Özlem SAKA

July, 2018, viii+94 pages

The purpose of this study is to unearth the role of LYS-5 (English Admission Test) English Test in predicting ELT students' language performance in four skills of the target language. For this purpose, predictive correlational research design of quantitative research methods is adopted in the study. The study group consists of 81 undergraduate students who are either newly-enrolled or studying in the department of English Language Teaching during the fall semester of the 2016-2017 academic year at a state university in a large province of southern Turkey. The data proceeded in this study are collected through Test of English Language Proficiency (TELP) which is a four-language-skill assessment test developed by the researcher and the collected data are analysed with the scores of the students' LYS-5 English Test by utilizing appropriate analysis techniques. The findings of the study reveal that LYS-5 English Test is a significant predictor of the students' success regarding the total scores of the students in the TELP; however, there is a low-level correlation between the scores of two tests. As a result of analysing the scores of each skill, it was found that although LYS-5 English Test meaningfully predicts the students' success in reading comprehension skills, the degree of the relationship is medium. Another prevailing finding of the present study is about listening, writing and speaking skills. The analysis results indicate that LYS-5 English Test is not a significant predictor of students' listening, writing and speaking skills. Besides, the results attest that the students attending to the extensive language training in the preparatory schools have higher scores than the ones who were exempt from the preparatory school or were not exempt from it but have not studied yet. In the lights of the findings, it is suggested that either the content of the current LYS-5 test should be extended in a way which will encompass four skills of the target or an alternative type of test should be developed in line with the requirement of the department of English Language Teaching.

Keywords: Undergraduate admission examination, LYS-5, testing, measurement and evaluation, university examination

ÖZET

İNGİLİZCE LİSANS YERLEŞTİRME SINAVI (LYS-5) İLE LİSANS ÖĞRENCİLERİNİN DÖRT DİL BECERİSİ ARASINDAKİ İLİŞKİNİN İNCELENMESİ

Yüksek Lisans, Yabancı Diller Eğitimi Anabilim Dalı

Danışman: Dr.Öğr. Üyesi Fatma Özlem SAKA

Temmuz, 2018, viii+94 sayfa

Bu çalışmanın amacı, LYS-5 (Lisans Yerleştirme Sınavı) İngilizce testinin İngilizce Öğretmenliği Bölümü öğrencilerinin hedef dilin dört becerisindeki dil becerilerini yordama durumunu ortaya çıkarmaktır. Bu amaçla, çalışmada nicel araştırma yöntemlerinden yordayıcı korelasyonel araştırma deseni kullanılmıştır. Çalışma grubunu Türkiye'nin güneyinde büyük bir şehirdeki devlet üniversitesinin İngiliz Dili Eğitimi Bölümü'ne yeni başlayan ya da hâlihazırda bu bölümde okuyan 81 İngiliz Dili Eğitimi öğrencisi oluşturmaktadır. Çalışmada kullanılan veriler, araştırmacı tarafından geliştirilen ve dört dil becerisini ölçen İngilizce Yeterlilik Testi (İYT) aracılığıyla toplanmış, toplanan veriler uygun analiz teknikleri kullanılarak öğrencilerin LYS-5 İngilizce testi sonuçlarıyla karşılaştırılarak analiz edilmiştir. Çalışmadan elde edilen bulgulara göre, LYS-5 İngilizce testi öğrencilerin İYT'deki toplam puanları açısından anlamlı bir yordayıcısıdır. Ancak, bu iki puan arasında düşük düzeyde bir ilişki vardır. Her bir beceriye ait puanların analizi sonucunda LYS-5 İngilizce testi öğrencilerin okuma becerilerindeki başarısını anlamlı bir şekilde açıklamasına rağmen arasındaki ilişki düzeyi orta olarak bulunmuştur. Bu çalışmanın diğer dikkat çeken bulgusu dinleme, yazma ve konuşma becerisiyle ilgilidir. Analiz sonuçları LYS-5 İngilizce testinin dinleme, yazma ve konuşma becerilerini anlamlı bir şekilde yordamadığını göstermektedir. Ayrıca, hazırlık programına katılan öğrencilerin hazırlık programından muaf olan ya da hazırlık programından muaf olmayıp henüz hazırlık eğitimini almamış öğrencilerden daha yüksek puanlar elde ettiği görülmüştür. Çalışmadan elde edilen bulgular ışığında ya LYS-5 İngilizce testinin içeriğinin hedef dilin dört becerisini kapsayacak şekilde genişletilmesi ya da İngiliz Dili Eğitimi Bölümünün gereklilikleri doğrultusunda alternatif bir test türünün geliştirilmesi önerilmektedir.

Anahtar Kelimeler: Lisans yerleştirme sınavı, LYS-5, sınav, ölçme ve değerlendirme, üniversite sınavı

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
ABSTRACT.....	ii
ÖZET.....	iii
TABLE OF CONTENTS.....	vi
LIST OF TABLES	vii
LIST OF ABBREVIATIONS	viii

CHAPTER I

INTRODUCTION

1.1. Background of the Study.....	1
1.2. Purpose of the Study	4
1.3. Significance of the Study	5
1.4. Scope of the Study	6
1.5. Limitation of the Study	6
1.6. Assumptions.....	6
1.7. Functional Definitions and Use of Concepts	7

CHAPTER II

LITERATURE REVIEW

2.1. Introduction.....	8
2.2. Assessment and Evaluation in Foreign Language Education	8
2.3. Types of Tests in Foreign Language Education.....	9
2.3.1 Achievement Tests	9
2.3.2. Language Aptitude Tests.....	10
2.3.4. Diagnostic Tests	10
2.3.5. Placement Tests.....	11
2.3.6. Proficiency Tests	11
2.4. Main Principles of Language Assessment	12
2.4.1. Reliability.....	12

2.4.2. Validity.....	14
2.4.3. Practicality.....	16
2.4.4. Washback.....	17
2.4.5. Authenticity.....	18
2.5. Competence vs. Performance.....	18
2.6. LYS-5 English Test.....	23
2.7. Preparatory Classes.....	24
2.8. Receptive and Productive Skills.....	27
2.9. Readiness.....	28
2.10. Related Studies.....	29
2.10.1. Washback Effects of Foreign Language Tests.....	29
2.10.2. Content and Changes in the Tests.....	32

CHAPTER III

METHODOLOGY

3.1. Introduction.....	36
3.2. Research Design.....	36
3.3. Study Group.....	36
3.4. Data Gathering Instruments.....	37
3.4.1. Proficiency Test (Test of English Language Proficiency).....	37
3.4.2. LYS-5 English Test Scores.....	46
3.5. Data Collection Procedure.....	47
3.6. Data Analysis.....	48

CHAPTER IV

FINDINGS

4.1. Introduction.....	50
4.2. Findings on the Values of LYS-5 Foreign Language Test in Predicting Four Language Skills.....	50
4.3. Findings on Reading Comprehension Scores of the Participants.....	52
4.4. Findings on Listening Comprehension Scores of the Participants.....	53
4.5. Findings on Writing Scores of the Participants.....	54

4.6. Findings on Speaking Scores of the Participants	55
4.7. Findings on Overall Test Scores of the Participants	56

CHAPTER V

CONCLUSION

5.1. Introduction	58
5.2. Results and Discussion.....	58
5.3. Suggestions for Practitioners and Policymakers	61
5.4. Recommendation for Further Research and Researchers	62
REFERENCES.....	63
APPENDICES	74
Appendix A: Test of English Language Proficiency	74
Appendix B / Scoring Rubric for pronunciation	87
Appendix C/ Scoring Rubric for intonation and stress	87
Appendix D/ Scoring Rubric for describing a picture	87
Appendix E/ Scoring Rubric for responding questions (3).....	88
Appendix F/ Scoring Rubric for expressing an opinion	89
Appendix G/ Scoring Rubric for summarizing a Text.....	90
Appendix J/ Scoring Rubric for writing an Essay.....	91
Appendix K/ LYS-5 English Test scores of the participants	92
CURRICULUM VITAE.....	93

LIST OF TABLES

Table 2.1. Communicative language competence in the CEFR.....	20
Table 2.2 Overview of perspectives on performance.....	23
Table 2.3 LYS-5 English Test Description.....	24
Table 2.4 Equivalence Table of Levels.....	25
Table 2.5 Teaching and Assessment Practices in Preparatory School.....	25
Table 3.1 TELP Questions Item Analysis Result.....	39
Table 3.2. Description of TELP Test.....	40
Table 3.3. Description of Reading Comprehension Section.....	41
Table 3.4. Description of Listening Comprehension Section.....	42
Table 3.5. Description of Writing Section.....	43
Table 3.6. Description of Speaking Section.....	45
Table 3.7. Inter-rater reliability results.....	46
Table 3 .8. Descriptive Statistics for Different Variables.....	48
Table 4.1 Model Summary and ANOVA Results.....	50
Table 4.2 Regression Model.....	51
Table 4.3 Multiple Relations CI Values.....	51
Table 4.4 Model Summary and ANOVA Results.....	51
Table 4.5 Regression Model.....	52
Table 4.6 Descriptive Statistics for Reading Section in TELP.....	53
Table 4.7. Descriptive Statistics for Listening Section in TELP.....	54
Table 4.8.Descriptive Statistics for Writing Section in TELP.....	55
Table 4.9. Descriptive Statistics for Speaking Section in TELP.....	56
Table 4.10. Descriptive Statistics for TELP Scores.....	57

LIST OF ABBREVIATIONS

CEFR: Common European Framework of Reference for Languages

CLT: Communicative Language Teaching

ELT: English Language Teaching

EPI: English Proficiency Index

FLE: Foreign Language Education

IELTS: International English Language Testing System

LYS-5: English Admission Test

MONE: Ministry of Education

PTE: Pearson Test of English

SPSS: Statistical Package for the Social Sciences

OSYM: Student Selection and Placement Centre

TELP: Test of English Language Proficiency

TOEFL: Test of English as a Foreign Language

YDS: Test of Foreign Language

YGS: University Entrance Examination

YOK: Council of Higher Education

YOKDIL: Foreign Language Test for Higher Education Institutions

YOS: International Student Exam

***KPDS, UDS, YDS and YOKDIL** are the high-stakes foreign language tests in Turkey. **KPDS** and **UDS** were collected under the heading of **YDS** which is used both in higher education, associate professorship, and academic staff applications and to get language allowance.

****YOKDIL** is used in higher education, associate professorship, and academic staff applications.

CHAPTER I

INTRODUCTION

1.1. Background of the Study

In today's world, learning a foreign language is considered as a prerequisite for accessing into a broader perspective other than one selves. Each passing day, the popularity of learning a foreign language is increasing and people are always in search of developing their skills in the target language. Although it was traditionally regarded as a "mimetic" activity, a process that covered students' repetition and imitation, foreign language learning and teaching practices underwent a significant paradigm shift by the effect of contemporary language learning theories (Moeller and Catalano, 2015). Throughout history, different language learning theories proposing that languages are learned through repetition, innate knowledge, interacting or connecting existing and new knowledge are put forward (Lightbown & Spada, 2013). Moreover, learning and teaching practices which were formerly implemented through teacher-centred learning environment, in which the teachers are the authority and students are the passive learners, began to be carried out in a student-oriented atmosphere that the teachers facilitate students' learning rather than controlling and the main focus became the students (Garrett, 2008). Nevertheless, the ever-growing desire to advance their communicative competences in foreign language has forced many countries to reshape their language policies in the last few decades.

Starting in the 1970s, Communicative Language Teaching (CLT) has served as a fundamental source in foreign language education practices which enables learners to master the target language with a high level of accuracy and fluency around the world (Richards, 2006). Madsen (1983) defined the era as "a time when we emphasize evaluation of language use rather than language form."(p.6). Unlike other instructional ways of teaching used in foreign language teaching, CLT doesn't build up a standardized framework, it is rather a set of principles which is mainly bounded to the individuals or institutions (Çelik, 2014).

From the very beginning of language learning and teaching, achievement in learning was in question. Thus, the ever-increasing interest in learning a foreign language has triggered the need for assessing language skills of the learners. As well as helping learners master the language, well-prepared language skills assessment also promotes learners' motivation (Madsen, 1983). Additionally, language assessment is broadly used in educational settings for a wide range of purposes such as placing students into the specific level or program, evaluating the efficiency of teachers or educational programs, selecting for employment or certifying individuals' language ability, and diagnosing learners' needs; besides, it constitutes a process in which the samples of learners' language performance are gathered, recorded, interpreted and some assumptions are made by relying on the interpretations (Bachman & Palmer, 2010). Moreover, among aforementioned functions, assessing language proficiency is predominantly considered as a common practice administered all around the world. Bearing in mind that a language consists of two principal phenomena which are labelled as receptive skills (reading and listening) and productive skills (writing and speaking) (Council of Europe, 2001), and language proficiency tests are generally centred upon those skills.

The recent research in testing skills (Hirai, 2002) points out that although learners noticeably master receptive skills which cover the understandings of the target language, the same success is not achieved in the productive skills which describe appropriate use and command of the language. Since there are some researchers (Brock, 1998; Gilfert, 1996) who have the opinion that the students' achievement in receptive skills is not a strong estimator for their achievement in productive skills, the studies conducted in this area aimed to explore the main points that will be used practically to advance their receptive and productive skills.

Since language assessment practices differ greatly in terms of the purpose of the test, the differences in the practice of internationally-accepted high-stakes tests such as TOEFL, TOIEC, IELTS, and PTE etc. are remarkable. For instance, administered as paper-based or internet-based, TOEFL measures candidates' general proficiency in the command of English and it is widely used as a criterion for international student admissions (Educational Testing Service, 2009). TOEIC, on one hand, measures daily-life language ability of a candidate working in an international workplace (GEPT

Research, 2002). Since it is divided into two areas such as academic and general, IELTS measures candidates' language skills based on the type of the test they attended (IELTS, 2018). Considering the varying purposes on the tests, some criteria concerning length and duration of the test, question type, test parts, number of questions etc. show differences across the tests.

In Turkey, language assessment practices are mainly carried out by OSYM (Student Selection and Placement Centre). Within the national standards, OSYM is in charge of preparing and administering two well-known language proficiency tests such as YDS and YDT (Test of Foreign Language), formerly known as LYS-5. While YDS is mainly administered as a test to measure state personnel's' language proficiency to receive language allowance (OSYM, 2018), LYS-5 is one of the higher education placement exams coordinated by OSYM (Council of Higher Education, 2017). Even though aforementioned tests such as TOEFL, TOIEC and IELTS measure candidates' language proficiency in four skills, YDS and LYS-5 both consist of 80 multiple-choice items which are primarily centred upon vocabulary, grammar, translation, and reading comprehension skills (Council of Higher Education, 2017). Unlike YDS, LYS-5 English Test serves as a compulsory test for assessing candidates' proficiency; therefore, it plays a significant role in depicting a detailed picture of candidates' language proficiency and placing them into the universities. Under these circumstances, LYS-5 English Test takes an active role in leading pre-service teachers and shaping the design of future teachers' community.

Considering that LYS-5 lacks assessing candidates' language proficiency in four skills, the candidates choose the departments without being aware of their level of language skills and their readiness for the requirements of the department. Since they constitute the group of future language teachers, foreign language education is also affected directly or indirectly (Karabulut, 2007). Although the students are supposed to be ready for the training in the departments by the lecturers, their inadequacy in the language skills of the target language makes things difficult and the syllabuses are not lectured as planned or students cannot benefit from the lectures as it is expected.

Taking the current place of foreign language learning and teaching in Turkey into consideration, the present study attempts to close the gap regarding the foreign language education by administering an alternative test to assess pre-service teachers'

language proficiency and contributes to the related literature in the given field. The motive and aspiration will be presented in the following section.

1.2. Purpose of the Study

The starting point of the present study was to seek an answer to the question whether ELT students are ready for the lectures in the medium of English in terms of their proficiency in language skills and to put forward some solution recommendations to advance the qualities of pre-service teachers of English by depicting their current situation and their readiness for the department of English Language Teaching. Therefore, the present study examines the role of LYS-5 English Test, which is the single scale assessing their proficiency in the target language, in predicting students' success in four skills of the target language.

For this purpose, the present study addresses the following research questions:

1. Does LYS-5 English Test successfully assess ELT' students' language skills?
 - a. Is LYS-5 English Test a significant predictor of students' success throughout the skills?
 - b. Is LYS-5 English Test a significant predictor of students' success in the reading comprehension skills of the target language?
 - c. Is LYS-5 English Test a significant predictor of students' success in the listening comprehension skills of the target language?
 - d. Is LYS-5 English Test a significant predictor of students' success in the writing skill of the target language?
 - e. Is LYS-5 English Test a significant predictor of students' success in the speaking skill of the target language?
2. Is there a significant relationship between the language proficiency test scores of the students who attended the preparatory school and the ones who did not?

1.3. Significance of the Study

English as a foreign language has long been taught in Turkey for many years. Although students start taking language classes from the early ages and it maintains throughout their education life, the results of foreign language learning are not satisfying (Habaci, Erdik, Adiguzelli & Sivri, 2011). Considering this issue and failed practices in foreign language learning, new foreign language teaching program has been implemented in accordance with the CEFR and agreed on an action-oriented approach in order to motivate learners to use English communicatively instead of focusing solely on the form (MONE, 2013). However, the output was not at the desired level. According to the 2017 report of English Proficiency Index, Turkey' proficiency was labelled as very low proficiency ranking 62nd of 80 countries (EPI, 2017)

All these results lead us to recheck the potential areas that urgently need further analysis and consideration. In the light of these practices, the present study examines the readiness of ELT students and assess their proficiency in English. It is a significant research since it both aims to unearth missing areas in language teaching and sheds lights on the studies in the related field and puts forward probable suggestions to improve the quality of pre-service teachers of English.

Although LYS-5 English Test serves as the single test that is administered by OSYM (Student Selection and Placement Centre) to place the students into the universities, alternative measurement and evaluation practices are suggested in the lights of the findings obtained from the present study. The present study might also contribute to the shift for the design of the current test and to uncover the missing parts that are supposed to assess other languages skills (listening, writing, and speaking) of the target language.

In addition, the present study compares the English proficiency test scores of the ones who attended an extensive English preparatory program and the ones who didn't. By this way, it highlights the significance and effectiveness of undergoing the preparatory school training in terms of language learning. Thus, the present study contributes concrete evidence to the debate on whether the preparatory schools should be optional or compulsory especially for ELT students.

Finally, the present study might contribute to the further research aiming at developing English pre-service teachers' proficiency and might provide a basis for prevailing factors.

1.4. Scope of the Study

The present study aims to ascertain the role of LYS-5 English Test in predicting ELT students' performance in four language skills of the target language. The study group consists of 81 undergraduate ELT students who are either newly-enrolled or studying in the Department of English Language Teaching during the 2016-2017 academic year at a state university in a large province in southern Turkey. This study attempted to unveil to what extent the participants were capable of using four skills of the target language, whether there is a significant difference between the language proficiency test scores of the ones who attended to the preparatory school training and the ones who didn't. Thus, the participants' readiness status has been specified in detail; some solutions and alternative measurement and evaluation practices were proposed to advance the current situation.

1.5. Limitation of the Study

Although the present study was carefully designed and conducted, it has some unavoidable limitations. Initially, it is limited to the study group of 81 undergraduate students studying in the Department of English Language Teaching during the fall semester of 2016-2017 academic year in a state university in Southern Turkey. It should also be highlighted that the findings obtained through the research were based on the participants' performance displayed in the test. Therefore, it was supposed that they fully understood the items in the TELP (Test of English Language Proficiency) and provided honest and fully-concentrated expressions of their knowledge.

1.6. Assumptions

While carrying out this research, the following assumptions are made. It is assumed that:

1. The TELP (Test of English Language Proficiency) test will cover the requirements of each language skill (reading, listening, writing, and speaking).
2. The TELP test will represent the characteristics of a B2 level language test.

1.7. Functional Definitions and Use of Concepts

Assessment: Although the terms assessment, evaluation and testing are interchangeably used in the educational settings, differences in the use of the terms have emerged in recent years. In this thesis, assessment is used to define the process of documenting information to monitor progress and make interpretations about learning (Overton, 2008). In addition, the terms assessment and testing are interchangeably used in this thesis.

Evaluation: The concept of evaluation does not merely describe the process of gathering numerical data from measurable attributes such as performance, skills, or attitudes, it also depicts the process to determine how well a student is successful and how much he/she has achieved the given subject (Bachman & Palmer, 2010). Therefore, evaluation is used in this thesis to refer to the process of making judgments based on the pre-set criteria or evidence to determine whether the given subject fulfil it or not.

Measurement: While it has a narrow-scope compared to the concept of evaluation, measurement refers to the process of exploring quantitative performance or abilities of the subjects and presenting the results through scores, percentage, and average (Wither, 2017). In measurement, a standardized tool is required to maintain the act of measuring. In line with the given definition, the concept of measurement in this thesis is used to refer to numerical data and standardized tests which are used to compare the success ratio of the subjects.

CHAPTER II

LITERATURE REVIEW

2.1. Introduction

In this chapter, the theoretical framework that underlines the research problem will be explained first. In this sense, the research will present a brief information about assessment practices and main test types in the foreign language, fundamental concerns of an assessment, theoretical basis of language proficiency regarding competence and performance, national examination of foreign language, preparatory classes, productive and receptive skills, and students' readiness. After the theoretical framework, within the scope of the literature review, the related studies are explained with regard to the research problem.

2.2. Assessment and Evaluation in Foreign Language Education

The need for assessment and evaluation has long been a significant issue since the early stages of the teaching process. Especially within teaching systems, it is requisite to measure the students' learning in order to guide them for an effective learning (Hughes, 2002). Even if it is commonly associated with the standardized proficiency tests in foreign language teaching, testing and evaluation encompass a vast range of practices concerning the purposes and forms of the tests. Along with these tests, teacher-made tests and other forms of tests such as Achievement Test, Language Aptitude Test, Diagnostic Test and Placement Test play a crucial role for an effective language teaching. Whilst the concept "assessment" is used as an "*umbrella term*" which embodies a series of methods and techniques (Biggs & Tang, 2007), it is often characterized as either "summative" or "formative" regarding the purpose for administering the assessment (Surgenor, 2010). Known as "classroom-based" assessment, the formative assessment refers to in-class observations and practices which are used to monitor students' learning (OECD, 2005). It furthers the teachers to guide and advance the learning and to achieve a comprehensible understanding of the learners by providing nonstop feedback (Bloom, Hastings & Madaus, 1971). Thus, the teachers can adopt different remedial methods to promote the learning. Since it lasts

during the entire learning process, it can be easily adapted to the classrooms in many ways (formal, informal, verbal, written etc.). Formative assessment is currently regarded as an integrated part of teaching-learning process rather than a separate evaluation process conducted at the end of the teaching (Allal, 1988; Audibert, 1980; Hamayan, 1995; Perrenoud, 1988). Summative assessment is a product-based assessment which is often used to evaluate the teaching process at the end of an instructional period (Anthony & Susan, 2005). In contrast to formative assessment, summative assessment is used as a means to evaluate the whole learning process at a particular time range. In this context, Stake (n.d.) describes the clear-cut distinction between formative and summative assessment through these words “*When the chef tastes the sauce it is formative assessment; when the customer tastes, it is summative.*”

2.3. Types of Tests in Foreign Language Education

Contrary to popular beliefs among some teachers, tests are not administered to catch students unready for assessing specific skills or attributes, or they are not developed to reinforce students to study. In contrast, the tests are considerably practical and a strong indicator for depicting the picture of learning, its environment, output, readiness, the efficiency of techniques and methods, and present condition. Since the tests are designed for various purposes, different types of tests are developed considering specific purposes. Although there are varying views on the types of tests, the tests are characteristically grouped under five headings in terms of their purposes as “Achievement test, Language Aptitude test, Diagnostic tests, Placement test, and Proficiency test” (Brown, 2004).

2.3.1 Achievement Tests

An achievement test is a test which is designed to measure how much an individual has learned within a specific program, a syllabus, or unit of instruction (Brown, 2004). Paterno (1965) characterized that “An achievement test is an inquiry to see if what has been taught is retained” (p.376). Brown (1994) further added, “tests that are limited to a particular material covered in a curriculum within a particular time frame” (p.259). In other words, an achievement test enables learners to perceive their individual progress in a specific curriculum within a fixed time. Since they are commonly used by language teachers to check whether the students have mastered the predefined

skills, what kinds of strengths or weakness the students have, how much learning-teaching process has achieved its objectives, to what extent techniques and methods used during the process were effective, they are sometimes called as “teacher-made test”. The achievement tests are usually summative since they are administered at the end of the teaching program to evaluate students’ learning (Brown, 2004; Rahman & Gautaman, 2012).

2.3.2. Language Aptitude Tests

Before moving the language aptitude tests, it is better to explain the concept of “Language Learning Aptitude”. Language learning aptitude is specified as the particular abilities that have an impact on language learning (Lightbown and Spada, 2006). Carroll (1991), one of the notable researchers in this field, defined it as quick-learning ability. Therefore, it can be assumed that “a learner with high aptitude may learn with greater ease and speed” (Lightbown and Spada, 2006, p.57).

Language aptitude tests became a crucial as a result of inadequate identification of learners’ language learning potentials through proficiency tests. Alderson and Hughes (1981) highlighted the reasons for the need for a language aptitude test:

“...a proficiency test, which simply labels a student at one point in time, gives no information about learning potential, and for that very reason may be inadequate. Two students may achieve the same proficiency score, but have very different potential: one student may have greater aptitude or adaptability than the other, perhaps having learnt the foreign language for only six months, whilst the other has studied it for fifteen years: in such a case one might expect the student with the shorter learning history to have greater potential for coping in a foreign language environment.” (p.125)

In other words, while the proficiency tests do not provide a detailed information about learning history or characteristics of examinees, the language aptitude tests aim to discover the reasons behind the score by viewing the learners’ characteristics.

2.3.4. Diagnostic Tests

The primary focus of this test is to find out the missing linguistic aspects to be taught during the course. Therefore, it plays a crucial role in organising an effective learning process and providing comprehensible input for missing skills. Rahman and Gutman (2012) stressed on the importance of employing a diagnostic test by stating that the teachers might be aware of the points to be emphasized and the skills to be practised

more through diagnostic tests. Bearing in mind that these tests reveal the gaps of both learners and educational resources, the teacher will adopt certain remedial methods to remove barriers in the learning process (Rahman & Gutman, 2012). Compared to other types of tests, the diagnostic tests are not either used in language teaching or preferred by teachers or institutions due to the difficulties in developing them; alternatively, achievement or placement tests are occasionally opted for diagnosing learners' strengths and weakness.

2.3.5. Placement Tests

As its name denotes, a placement test is primarily designed to provide useful information for placing students into an appropriate level of a program or a course in line with their scores. This test is mostly used to decide which group or class the students should be assigned (Hughes, 1996). Moreover, the placement tests are also beneficial for teachers since the students are placed into the appropriate levels and so the teachers might be aware of the language abilities of the target group (Alderson, 1996). From this perspective, placement tests and diagnostic tests might seem relatively similar; however, while diagnostic tests are used to get the picture of students' knowledge about a language, the placement tests are applied to place them into the groups or levels by utilising the existing knowledge of the students in a language (Ozerova, 2004). To put it in a different way, diagnostic tests focus on the knowledge of the individuals about a language and tries to unearth missing skills of that language; the placement tests, on the other hand, center upon the outcome obtained through existing knowledge of the individuals about a language to assign them into the appropriate levels of groups or classes. Since they are designed depending on the purpose of the institution and the syllabus of the course, any and every institution has its own placement test, so it is possible to encounter several kinds of placement tests in language teaching (Hughes, 1996). Nowadays, the placement tests are largely administered by language institutions to place the learners into the levels and to create homogenous classroom settings.

2.3.6. Proficiency Tests

Proficiency tests are not limited to a particular program or unit, but measures individuals' general level of mastery in language regardless of any training that they

may have received in the target language (Brown, 2004; Huges, 1996). Valette (1997) further adds that “the aim of a proficiency test is to determine whether this language ability corresponds to specific language requirements.” (p.6). In a similar manner, Hamp-Lyons (1998) describes the proficiency tests used to measure individuals’ readiness for a specific communicative competence in that language. Some researchers state that these tests provide an overall measure of “language ability” (Deshang & Varghese, 2013; Hamp-Lyons, 1998) and are also known as “high-stakes tests” in the literature (Desheng & Varghese, 2013). Considering that these tests are limited to a specific content or objectives of a course and an individual who achieves predefined specifications is considered to be proficient in that language, the researchers seek an answer to find out what is associated with the word “proficient” (Hughes, 1996). Regarding the lexical meaning, the word “proficient” is defined as “able to do something well or skilfully.” (Longman Dictionary of Contemporary English, 2018), “competent or skilled in doing or using something” (Oxford Living Dictionaries, 2018), or “skilled or experienced” (Cambridge Dictionary, 2018). The word “proficient” is described as having sufficient command (mastery) of a language in terms of a particular purpose in some proficiency tests (Hughes, 2002). Therefore, as Brown (2004) noted, specifying the constructs of language ability has been the fundamental issue in assessing language proficiency. Although there is an on-going debate on the construct validity of proficiency test, some standardized proficiency tests such as TOEFL, TOIEC, IELTS, PTE, and Cambridge English exams are officially recognized by colleges and institutions, so every year thousands of examinees take these exams for different purposes such as job/visa applications, college submissions, working or living in English-speaking countries. Even if the proficiency tests differ in terms of content and level of difficulty, the only thing they have in common is that they do not focus on a course which the examinees may have formerly taken (Hughes, 2002).

2.4. Main Principles of Language Assessment

2.4.1. Reliability

It is the degree of the consistency between two treatments of a test with the same students or bi-level students. Madsen (1983) defines a reliable test “the one that produces essentially the same results consistently on different occasions when the

conditions of the test remain the same” (p.179) However, since he believes that people never act in the same way on every occasion, Hughes (1889) considers that attaining a hundred percent of reliability is nearly impossible even when all the aspects of a test is identical. Bachman (1990) states that factors such as test environment, characteristics and condition of the participants (age, gender, fatigue, illness etc.) have a significant impact on whether the test is completely reliable or not.

Hughes (1989), however, points out a linear relationship between the number of the test items and the degree of the reliability. In other words, the more items a test has, the higher reliability can be achieved. A very crucial criterion of this condition is to make additional items independent from one another; otherwise, the learners are exposed to double-punishment when they responded to the original question incorrectly (ibid.).

The concept of reliability is often associated with the validity. Hughes (2003) highlights that “if a test is not reliable, it cannot be valid” (p.34) and adds “to be valid a test must provide consistently accurate measurements” (p.50). In that vein, Guilford (1990) approaches the relationship between two dimensions and specifies that while the reliability of a test can be established without being valid, it is not possible to achieve a valid test without being reliable. However, Heaton (1988) disagrees with these ideas by claiming that: “The greater the reliability of a test, the less validity it usually has” (p.164).

The reliability is mainly discussed under four subheadings (student-related, rater, test-administration, and test reliability).

2.4.1.1. Student-related Reliability

It encompasses the factors emerging from the participants such as illness, bad day, fatigue, anxiety, nervousness, and other physical and psychological factors (Mousavi, 2002). Although the participants have a mastery of behaviour domain measured in the test, they may not display their actual performances when they are faced with any kind of those factors in the test.

2.4.1.2. Rater Reliability

Another crucial aspect of reliability is the rater reliability also known as “scorer reliability”. It primarily refers to the degree of agreement between the raters (Alderson,

Clapham & Wall, 1995). The factors such as subjectivity, human errors and bias can opt into the scoring process (Brown, 2004) and they may mislead test scores by decreasing the reliability of the test. This form of reliability is also categorised as inter-rater reliability, which is the balance/agreement between scorings of two or more raters, and intra-rater reliability, which is the degree of agreement between two recurrent administrations of a test scored by a single rater (Alderson et al., 1995). To overcome rater-related reliability issues, Linville (2011) suggests that especially for subjective and open-ended assessments, outlining the scoring process in advance and training the scorers will improve the balance and enable appropriate assessment.

2.4.1.3. Test Administration Reliability

It covers the conditions of test administration such as noise, the height of tables, temperature and lightness of classroom, the inefficacy of materials and other physical and environmental factors (Brown 2004).

2.4.1.4. Test Reliability

Another form of reliability is the test reliability of the test itself. A long test can cause fatigue for the test-takers and they may swiftly respond the items in order to finish up the test on time (Brown, 2004). Time-scaled tests can hinder some test-takers' performances and even though they have a good mastery of a specific course, their test scores do not adjust with the reality (Hughes, 2003). Besides, such factors as poorly written items and bad quality of the copies may affect test reliability adversely (Wells, 2003).

2.4.2. Validity

The concept of validity deals with whether a test precisely measures what is expected to measure (Hughes, 1989). Gronlund (1998) characterized it as "the extent to which inferences made from assessment result are appropriate, meaningful, and useful in terms of the purpose of the assessment" (p.226). In establishing a valid test, the purpose should be on the intended attribute. In other words, if the concern of an assessment to test students' speaking skills, the test should be solely constructed in a way which merely assesses speaking (Linville, 2011). Although there is no ultimate,

definite measure of validity, a validity of a test can be ascertained by providing several kinds of validity evidence regarding the test (Brown, 2004).

2.4.2.1. Content-related Evidence

It refers to what extent the given test encompasses a representative and relevant sample of the relevant subject which is intended to measure (Hughes, 1989). In a similar manner, Anastasi (1982) described it as “essentially the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured.” (p.131).

In case of attaining a significant content validity, a test can truly measure what is really supposed to measure. Whilst it may be difficult to establish content validity, a number of guidelines can be followed:

1. the behaviour domain to be tested must be systematically analysed to make certain that all major aspects are covered by the test items, and in the correct proportions’;
2. the domain under consideration should be fully described in advance, rather than being defined after the test has been prepared;
3. content validity depends on the relevance of the individual’s test responses to the behaviour area under consideration, rather than on the apparent relevance of item content (Anastasi, 1982).

2.4.2.2. Criterion-related Evidence

Another form of evidence of the validity of a test is criterion-related evidence (validity). It describes how the test scores and external performance outside the test are correlated with one another (Weir, 1990). It is also grouped under two subheadings: concurrent and predictive validity. While concurrent validity relates to the existing performance, the predictive validity deals with how successfully the test scores predict test-takers’ future performance (Bachman & Palmer, 1981).

2.4.2.3. Construct-related Evidence

Construct-related evidence can be described as the degree of a test in measuring an underlying ability (trait) to be supposed to test (Hughes, 1989). Construct validity is generally established by comparing the test which is under consideration with other

tests that measure similar behaviour domains and checking how significantly correlated they are. The construct-related evidence is the broadest among all the forms of validity evidence. Since the large-scale tests of proficiency should consider practicality and measure test-takers' language proficiency through a limited number of language areas, construct validity is considered as a key issue for the validity of those tests (Brown, 2004).

2.4.2.4. Consequential Validity

Consequential validity covers all the potential aftereffects (positive/negative-intended/unintended) of a particular assessment (Brown, 2004). Weir (1990) names it as "washback validity" and concentrates on the test in terms of particular uses, its effect on test-takers, social outcomes, and consequential evidence for teachers.

2.4.2.5 Face Validity

"Face validity refers to the degree to which a test looks right, and appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of the examinees who take it, the administrative personnel who decide on its use, and other psychometrically unsophisticated observers" (Mousavi, 2002, p. 244). When one can understand what is expected to measure by looking at the test, it means the given test has face validity (Patton, 1997). On the one hand, the concept of face validity is considered significant for maintaining test-takers' motivation and performance (Heaton, 1975; Weir, 1990). On the other hand, since the judgment is subjective, there is an ongoing debate on the usefulness of this validity (Roberts, 2000).

2.4.3. Practicality

The final characteristic of a good test is its practicality. According to Hughes (1989), even "the best test may give unreliable and invalid results if it is not well administered" (p.152). In case of its inefficiency concerning economy, administration, scoring and interpretation, a test having high validity and reliability evidence can easily turn to a poor one (Weir, 1990) The concept of practicality is generally associated with the evaluation of the assessment with regard to cost, required time, and usefulness (Yoneda, 2012). The characteristics of a practical test are explained by Brown and Abeywickrama (2010) as follows:

A practical test:

- stays within budgetary limits
- can be completed by the test-taker within appropriate time constraints
- has clear directions for administration
- appropriately utilizes available human resources
- does not exceed available material resources
- considers the time and effort involved for both design and scoring

(p. 26)

2.4.4. Washback

Testing is not regarded as a stable process; on the contrary, it always has consequences. The effect of testing on learning and teaching is called “washback”, also known as “backwash” (Alderson & Wall, 1993; Biggs, 1995; Cheng & Curtis, 2004; Hughes, 2003; Saville, 2000). Shohamy, Donitsa-Schmidt, and Ferman (1996) interpret in a simple manner as "the connections between testing and learning." (p.298). In a broader sense, Messick (1996) sheds light on the definition of the concept of washback that "washback, a concept prominent in applied linguistics, refers to the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning." (p.241). As well as the varying definitions of the concept, the direction of washback is described as positive or negative (Bachman & Palmer, 1996; Davies et al., 1999; Heaton, 1990; Hughes, 1989; Messick, 1996), intended or unintended (Andrews, 2004; Qi, 2005)

This term is commonly used as a synonym for ‘effect’, ‘impact’, ‘consequence’ and ‘outcome’ in the field of language assessment. Especially for large-scale assessment, washback usually refers to the effects on the teaching of the behaviour domain and students’ preparation for the test (Brown, 2003).

Messick (1989, 1996) puts forward a relationship between washback and consequential validity and his idea is also supported by other researchers (Frederiksen & Collins, 1989; Shohamy et al., 1996; Weir, 1990) Messick (1996) places it within the concept of consequential validity and states that as a part of unified validity, washback is a proof of the consequential facet of validity, which is merely one of the

six essential forms of evidence leading the validity of interpretation and treatment of the language.

2.4.5. Authenticity

Although it is a debatable concept among the researchers, the concept of authenticity is one of the forms of the reliability (Wood, 1993) and considered as one of the crucial factors of a good test (Bachman & Palmer, 1996; Carroll, 1980; Weir, 2005). Bachman and Palmer (1996) characterize it as “the degree of correspondence of the characteristics of a given language test task to the features of a target language task,” (p.23). It, in the simplest way, is the match between the language of test items and the target language. Therefore, authentic tests are often regarded as the simulations of real-life language tasks (Shomoossi, 2010). They are referred to “communicative”, “direct”, “performance” tests etc. (Jian-lan, 2007). In addition to the supporters of the concept, some opposing views have aroused since it was first introduced to the educational paradigm. Lewkowicz (1997) sees it as a relative concept and adds that while some may perceive an item as authentic, others may consider it “non-authentic”.

2.5. Competence vs. Performance

Even though there are terminological misconceptions and debates concerning the concept of “competence”, it is commonly used in linguistics and language teaching, especially in generative grammar proposed by Noam Chomsky, the pioneering innateness linguist. Chomsky (1965) defined it as “the speaker-hearer's knowledge of his language.” (p.4). In his definition, Chomsky made a fundamental distinction between competence and performance. As stated, he described competence as an innate capability to acquire one’s native language and distinguished from performance (actual ability to use a language). Upon this definitions, some certain interpretations have been made by linguists and researchers. McNeil (1966) defines competence as “the knowledge of linguistic rules, categories, etc., that accounts for native speaker’s intuitions about his language; the expression of such competence in talking and listening is performance.” (p.77). W. White (1959) describes competence as “an effective interaction of the individual with the environment”. Weinert (2001) touches upon the ongoing issue and postulates that some certain dimensions have a significant impact on a person’s degree of competence and lists the following figures:

- Ability
- Knowledge
- Understanding
- Skill
- Action
- Experience
- Motivation (p.29)

Over the years, definitions of competence have shown a consistent progression from the first proposed by Chomsky. Based on the current interpretations of the term, Newby (2011) gathers three prominent hypotheses below:

“The first is the widespread acceptance that language is essentially a cognitive phenomenon and that the use of the linguistic code of a language (performance) is steered by tacit rule-based knowledge stored in the minds of speakers (competence). This view includes both a Chomskyan modular (language-specific) view of competence and those theories that can be grouped under the heading of ‘Cognitive Linguistics;’ which see language and cognition in general as an integrated whole sharing similar systems of perception and categorisation.” (p.15)

Chomsky used “linguistic competence” to refer to a person’s linguistic knowledge. Even though this competence was considered as a fundamental entity that could be best described regarding grammatical rules of a language, new types of competences have been proposed by opponent linguists. Hymes (1972) introduced “Communicative Competence” as a reaction to linguistic competence. This type of competence deals with individuals’ ability to use the language in a variety of authentic situations as well as having an inherent grammatical competence. Adding to Hymes’ definition, Widdowson (1983) clarified the concept of communicative competence by dividing into two distinctive notions (competence and capacity). While he described competence as a person’s knowledge of linguistic and sociolinguistic systems, he referred capacity as the ability to use existing knowledge as a way of having comprehensible interactions in the given language. Canale and Swain (1980) and Canale (1983) figure out that communicative competence requires the combination of existing knowledge about a language and an individual’ skills for communications. Therefore, they expanded the notions of communicative competence. According to Canale and Swain (1980), communicative competence arises from three components:

1. Grammatical Competence: knowledge of lexical items and rules

2. Sociolinguistic Competence: coherence and rules of discourse
3. Strategic Competence: use of communication strategies

Since communicative language teaching practices are predominantly implemented in language learning and teaching, communicative competence primarily arouses researchers' attention. Those skills have also been integrated into the CEFR (Common European Framework of Reference for Languages). In CEFR, communicative language competencies are grouped into three categories "Linguistic Competences, Sociolinguistic Competences, and Pragmatic Competences" and those categories are presented below:

Table 2.1. Communicative Language Competence in the CEFR

Communicative language competence (CEFR 2.1.2, p.13)	
Linguistic competences	'lexical, phonological, syntactic knowledge and skills and other dimensions of language as system' (13).
Lexical competence (CEFR 5.2.1.1)	'knowledge of, and ability to use, the vocabulary of a language, consists of lexical elements and grammatical elements' (110)
Grammatical competence (CEFR 5.2.1.2)	'knowledge of, and ability to use, the grammatical resources of a language' (112)
Phonological competence (CEFR 5.2.1.4)	'a knowledge of, and skill in the perception and production of: the sound-units (<i>phonemes</i>) of the language and their realisation in particular contexts (<i>allophones</i>), etc. (116)
Orthographic competence (CEFR 5.2.1.5)	'a knowledge of and skill in the perception and production of the symbols of which written texts are composed' (117)
Sociolinguistic competences	'refer to the sociocultural conditions of language use' (...) 'rules of politeness, norms governing behaviour between generations, sexes, classes and social groups, linguistic codification of certain fundamental rituals in the functioning of a community' (13)
Pragmatic competences:	
Functional competences, (p.123)	'relating to the communicative function of utterances' (production of language functions, speech acts)
Discourse competences	'the mastery of discourse, cohesion and coherence, the identification of text types and forms' (13) 'relating to the organising and structuring of texts' (...) 'drawing on scenarios or scripts of interactional exchanges' (123).

By integrating skills, CEFR goes beyond from being solely knowledge of a language to the reflection of an individual's potential.

Especially with the innateness perspective which hypothesized the existence of a pre-programmed device “Language Acquisition Device”, endowed with Universal Grammar defined as the “system of categories, mechanisms and constraints shared by all human languages and considered to be innate” (Chomsky 1986, p. 3; O’Grady et al. 1996, p.734 ; Pesetsky 1999, p. 476), language acquisition underwent a fundamental distinction between competence “the speaker-hearer's knowledge of his language” and the performance “actual use of language in concrete situations” (Chomsky, 1965,p.4). The linguistics who studied second language acquisition from the innateness perspective plugged away at whether there is a significant relationship between the competence that stresses the performance or actual use of the language and the competence underlying linguistic performance of the speakers (Lightbown & Spada, 2006).

Whereas the concept of performance was interchangeably used in many disciplines, it became a separate notion with the distinction between Chomsky’s competence and performance. Compared to the concept of competence, the performance represents an individual’s actual use of a language to make meaningful interactions. Whilst Chomsky claimed that the performance is the fallacious indicator of competence and it does not comprise a tangible discipline, it was within Hymes’s (1972) field of interest since it is the product of “social interaction” (272).

Hymes (1972) attempted to define the concept of performance and described it as the following:

When one speaks of performance, then, does one mean the behavioural data of speech? or all that underlies speech beyond the grammatical? or both? (...)
The difficulty can be put in terms of the two contrasts that usage manifests:

1. (underlying) competence v. (actual) performance;
2. (underlying) grammatical competence v. (underlying) models/rules of performance (p.281)

Regarding the definitions above, Newby (2011) claimed that there is an ambiguity in the different senses of the term and especially the latter sense is compared with both grammatical and communicative competence. By defining the performance as a “behavioural potential”, Halliday (1978, p.38) made a definite interpretation of the concept. He stated that the individuals’ can-do statements refer to their behavioural potential as an indicator that might be enciphered in language or other things. In a broader sense, Hymes (1972) described that “It [performance] takes into account the interaction between competence (knowledge,

ability for use), the competence of others, and the cybernetic and emergent properties of events themselves” (p. 283). When we take the terms “interaction” and “cybernetic”, we can clearly understand that as Halliday stated, the performance does not merely represent behavioural potential; rather, it is the actual use of that language. The intensive analysis of the performance in terms of usage of definitions and interpretations reveals that the following three usages of the term are commonly preferred:

- a) the “behavioural potential” to perform (knowledge-based grammatical competence, pragmatic competence, discourse competence, etc.)
- b) the act of performing
- c) the product of performance (output)

(Newby, 2011, p.20)

Thanks to the efforts of the researchers, we have witnessed a progression in clarifying and extending the concept of the performance during the last decades (Campbell, 1990). Researchers have brought different perspectives in order to study performance. In the most general sense, they have adopted three different perspectives; individual differences perspective, situational perspective, and performance regulation perspective.

As stated in Table 2.2., the first perspective pays attention to the underlying individual factors (mental ability, personality, age, sex etc.) that cause performance differences among them. This perspective asserts the core question “Which individuals perform best?” and tries to enlighten performance differences through individual differences in abilities, personality and/or motivation (Sonntag & Frese, 2005). The second perspective (situational perspective) refers to the factors in individuals’ environments which reinforce, promote or burden the performance of the individuals. The primary focus is to find out in which situations the individuals perform best. Unlike the others, the final perspective (performance regulation perspective) do not take the individual or situational factors into account; rather, it focuses on the performance itself. By this way, it aims to conceptualize the function and manner of work of the performance.

Table 2.2 Overview of Perspectives on Performance

	Individual differences perspective	Situational perspective	Performance regulation perspective
Core question	Which individuals perform best?	In which situations do individuals perform best?	How does the performance process look like? What is happening when someone is 'performing'?
Core assumptions and findings	Cognitive ability Motivation and Personality Professional experience	Job characteristics Role stressors Situational constraints	Action process factors Adequate hierarchical level
Practical implications for performance improvement	Training Personnel selection Exposure to specific experiences	Job design	Goal setting Feedback interventions Behaviour modification Improvement of action process Training Job design

(Sonnentag & Frese, 2005, p.9)

While the phenomenon “performance” can be represented by these three perspectives and the differences among them are highlighted, many researchers (Barrick & Mount, 1993; Colarelli, Dean, & Konstans, 1987; Hackman & Oldham, 1976; Mitchell, 1997) choose to combine one or two perspectives when defining it.

2.6. LYS-5 English Test

Formerly known as LYS5, LYS English Test is one of the tests that constitutes the second phase of two-stage university entrance examinations. The students, getting the passing grade in the first examination entitled with YGS (University Entrance Examination) encompassing the expected knowledge, skills and competencies that candidates should possess in quantitative and verbal fields, gain the right to take LYS5 English Test. The test is administered by OSYM and approximately 40-50 thousands

of candidates take the test every year. It lasts 120 minutes and consists of 80 questions shown in the following table;

Table 2.3 *LYS-5 English Test Description*

Parts	N of Questions	Question type
Vocabulary	5	Multiple-Choice
Grammar	10	
Cloze Test	5	
Sentence Completion	8	
English-Turkish Translation	6	
Turkish-English Translation	6	
Reading Comprehension	15	
English-English Translation	5	
Paragraph Completion	5	
Situation Questions	5	
Dialogue Completion	5	
Odd One Out Questions	5	

2.7. Preparatory Classes

As an introductory program, the preparatory program helps students to prepare for studying in higher education. It covers an extensive foreign language training procedure in line with the objectives of the academic programs. The students enrolling in the universities take a proficiency exam and the ones who have an adequate score, are exempted and go directly to the academic programs in the faculties. The others undergo an extensive foreign language training for a year. Although it is practised in several languages, in most universities, English as a Foreign Language (EFL) constitutes a large part of the preparatory program. Nevertheless, course hours, the number of exams, quizzes etc. change across the universities.

The students are expected to be B2 level of the CEFR at the end of the program. Therefore, the exemption procedures are based on the equivalence table of Foreign Language Education and Examination Regulations of the universities.

Table 2.4 Equivalence Table of Levels

CEFR Levels	YDS/KPDS/UDS
A1	30
A2	45
B1	60
B2	75
C1	90

As presented in Table 2.4, the students who enrolled in diploma programs in the field of language are required to score 75 points or more from the national language tests to be accepted at level B2 by the majority of the universities, so they might be exempt from the compulsory preparatory program. The others start their training in the program. Especially for the preparatory school program in the university where the current research is conducted, the program covers 20 hours of training in a week. The courses are lectured in two semesters, fall and spring, which cover 14 weeks of teaching and learning activities (School of Foreign Language, 2017). It is held in three programs and the content of each program is illustrated in the table below.

Table 2.5 Teaching and Assessment Practices in Preparatory School

Program	Modules	Module Tests	Quizzes	Portfolio Assessments
P1	3	3	9	10
P2	3	3	9	9
P3	2	2	6	10

Based on their scores in the proficiency test administered by the university, the students are placed into the relevant program and continue their training according to the requirements of that program. While the first two programs consist of three modules and the students are exposed to nearly the same number of quiz, module test and portfolio assessment, the program 3 is carried out in two modules and the students in this program have less number of quiz and module test compared to the others. It is obligatory for students to continue to 85% of the courses in the preparatory program in order to attend the proficiency test at the end of the program.

While these programs are administered electively or compulsorily based on the requirements of the relevant departments by some universities, they play a crucial role in preparing students and developing their language skills for the departments where the medium of instruction is in English. Even if the medium of instruction is not officially in English for all the universities in Turkey, the courses; except for some basic ones, are inherently lectured in the target especially in the departments of Foreign Language Teacher Education (e.g. English Language Teaching, German Language Teaching, French Language Teaching etc.). Under these circumstances, the students' readiness in terms of language skills becomes even more critical in order to be able to follow the courses in their departments.

Today, an increasing number of universities have started to integrate English as a preparatory course to improve students' foreign language skills. In some universities, it is even offered as a medium of instruction for several courses. Despite the efforts and incentives of the universities, the results are not satisfying. The researches in the related area (Alici,2004; Karataş & Fer, 2011; Vural,2004) show that the students still endeavour to advance their foreign language proficiency even after they have completed the preparatory classes or graduated from the university. The current picture of language proficiency is also attested by English Proficiency Index in which Turkey is located in the 62th out of 80 countries and is categorised as very low proficiency (EPI, 2017).

2.8. Receptive and Productive Skills

Defined by Allen (1974) as “a means of communicating thoughts”, language occupies a crucial position in humans’ life. Whilst it encompasses a wide range of components, a language mainly consists of four basic skills which are grouped under two headings: Receptive Skills (Passive Skills) and Productive Skills (Active Skills) (Husain, 2015). In a broader sense, Aydogan and Akbarov (2014) group those language skills based on two factors; the way of interaction: oral or written and the direction of interaction: receiving or producing the message. Sadiku (2015), on the other hand, describes speaking and listening as highly interrelated skills to achieve an effective oral communication and underlines the strong relationship between reading and writing for gaining an effective written communication. Although the mastery of those skills follows a gradual process, learners will be able to communicate in the target language effectively only when they successfully acquire and integrate the aforementioned skills with one another. Harmer (2007) points out that when people interact with each other, any of the four fundamental skills are not practised separately, and rather they are correlated during the conversation. Hinkel (2006) specifies the significance of language skills’ integration by stating that “In meaningful communication, people employ incremental language skills not in isolation, but in tandem.” Oxford (2001) remarks that Integrative-Skill Approach, a recent approach aiming at advancing learners’ four language skills interdependently, ensures learners to have an ideal ESL/EFL communication.

While the maximum integration of four skills in learning and teaching is considered as a prerequisite for optimum communication, the practices of assessing those skills remarkably differ. Schellekens (2001) stresses that the learners’ language ability to speak and understand considerably differs when they are not at the expected norm and therefore the traditional understanding of assessment measures these skills separately. He also highlights a possible result in the event of assessing speaking and listening skills through these words:

“Candidates may have understood a dialogue or a question but may lack the language to report back or respond appropriately. Thus the listening score is ‘polluted’ by the candidate’s lower speaking skills. As a result, the speaking score may be accurate but the listening skills are rated below the actual level.” (p.11)

2.9. Readiness

The concept of readiness, which is highly important for the educational process, is the critical input of learning and teaching system (Bloom, 1995). Jensen (1969) characterises it as “the amount of previous learning that can transfer to new learning” (p.1). Downing and Thackary (1971) consider readiness as a period in which a child can easily learn a language without getting into any kind of difficulty. The well-known researcher Thorndike (1977) associated readiness with the condition of preparation. In other words, preparation and readiness are inter-correlated terms in educational settings and depending on the condition of being prepared/unprepared, conducting a new learning may be either satisfying or annoying. Concerning these views, it can be assumed that acquiring a new behavioural change in education depends on the level of readiness. Therefore, the students should have the prerequisite cognitive, emotional and motional behaviours (Başar, 2001).

Readiness level can be defined as the student’s background and it includes his/her general and specific skills, quality of his/her knowledge and skills for the target behaviours, interests, attitude and motivation (Bloom, 1995).

The items that shape the readiness encompass the individual's attitude towards learning, the concepts and values developed, self-esteem, habits, language development, interests, needs, methods and fears (Başaran, 1998).

Ünal (2005) grouped the concept of readiness under three subcategories (cognitive readiness, social readiness and educational readiness). Cognitive readiness refers to have the required ability and essential behaviours relating to learning (Tuna & Kaçar, 2005). Social readiness mainly focuses on the social environment where new learning takes place (Harman & Çelikler, 2012). In social readiness; while positive and supportive behaviours strengthen and render permanent learning, preventive and humiliating behaviours impede learning and give the sense of failure.

Since education is a process of behavioural change, it is highly important to measure the difference between input and output; thus, specifying the level of readiness is imperative to know where to start (Fidan, 1986).

The level of readiness is primarily determined:

- to find out their existing knowledge in the very beginning of the subject

- to explore students' individual characteristics,
- to decide on whether the previous learnings will be repeated or new learning will occur (Yüksel, Marangoz, and Canaran, 2004).

Considering the place of readiness in the educational process, several tests are administered to find out how well the students are prepared for a planned learning unit in terms of the level of knowledge and skills. Thanks to these tests, missing aspects are specified and achieved before going into the target learning system.

The students with high readiness grasp the subjects quickly, comment on them, make their assignments effortlessly since they have learned the concepts relating to the topics, and become ready for the following learning because they have understood the previous subject well (Yenilmez & Kakmacı, 2008).

2.10. Related Studies

When the related studies are reviewed, it is probable to come across a limited range of studies on foreign language testing, which leads us to rethink the significance of the present study. The studies on foreign language tests mainly focused on the washback effects (Çakıldere, 2013; Dağtan, 2012; Hatipoğlu, 2016; Karabulut, 2007; Sayın & Aslan, 2016; Sevimli, 2007; Yıldırım, 2010), content and changes (Arslan, 2005; Aşkaroğlu, 2014; Biltekin, 2004; Elaldı, 2005) in the structure of the tests over the years.

2.10.1. Washback Effects of Foreign Language Tests

Sevimli (2007) focused on the washback effects of LYS-5 English Test on learning and teaching procedures in high schools. The study was conducted with 13 teachers of English, 87 LYS-5 English Test candidates studying in the 10th or 11th grades in 6 high schools. During the research, in-class activities were observed, interviews were made with teachers, and the questionnaires were given to the teachers and students to explore the participants' views on in-class activities and LYS-5 English Test. The analysis results unveiled that LYS-5 English Test had a negative impact on learning and teaching practices in high schools. The curriculum is generally ignored and test-oriented programs were adopted by the teachers to help the students to succeed in the test. Instead of the resources based on the scientific research, the ones which focused

on the content of the test were used during the process. While the students' critical thinking skills were ignored, their memorization skills were fostered. The students did not practice such language skills (listening, writing, and speaking) since they were not considered as a prerequisite for succeeding in the test. The findings of this study are also supported by another study which was conducted by Karabulut (2007) and precipitated that except for reading comprehension skills, the others are neglected and the language courses are mainly covered with the activities of vocabulary, grammar, reading in high schools.

In the same vein, Sayın and Aslan (2016) conducted a study to find out the perceptions of undergraduate ELT freshman students on LYS-5 English Test. They collected the data from 74 students studying in two different universities located in the province of Samsun via questionnaires consisting YES-NO questions, checklists, and 5 point-Likert scale questions. The analysis results revealed that LYS-5 English test considerably affected their training in high school in terms of in-class activities. The students stated that in-class activities were primarily on reading, vocabulary practice, practising test-based questions and general assessment test. However, the self-report data obtained from the students unveiled that the activities related to neglected skills (writing, listening, and speaking) were not commonly practised in the classrooms. Besides, some students highlighted that they had difficulty following the courses in the faculties since they were not as proficient as they were supposed to be. However, they were aware of their current condition in the target language and demanded that the neglected skills in LYS-5 English Test should also be integrated since it will help them change their training in high school and develop their readiness for training in the faculties.

A similar study was carried out by Yıldırım (2010), though. In his study, he adopted a mixed method research design and collected the data from 6 ELT lecturers and 79 ELT students at a state university. The findings of the study showed similarities with the findings of the studies conducted by Sevimli (2007) and Sayın and Aslan (2016) who emphasized the negative washback effect of LYS-5 English Test on classroom activities and found out that the test in question merely assessed reading skill, knowledge of vocabulary and grammar; however, it did not measure listening, writing, and speaking.

Hatipoğlu (2016) shed light on existing findings regarding the washback effects of LYS-5 English Test from the perspectives of senior ELT students studying at a state university. By using survey questionnaires and semi-structured interviews, she collected the data proceeded in the study. The findings of the study attested that although LYS-5 English Test had a detrimental effect on the learning and teaching of English in Turkey, it did not serve as a practical tool for communication. Because of the content of the test, the students could not transfer their knowledge into real-life situations, had difficulty in identifying smaller units of the target language, and could not improve their competence in listening, writing, and speaking. Their progress in English was not measured and followed in the current system, either. She also pointed out the need for preparing a valid test which will have positive washback effect for students, learning and teaching of the target language, and the teachers.

Since the studies on LYS-5 English Test were confined to the above-mentioned studies, the researcher involved other studies which examined national language tests such as UDS, KPDS and YDS. While they serve at cross purposes, the content of the national language tests is nearly the same as LYS-5 English Test. Therefore, the findings of the available studies on these tests can also enlighten the results of the present study.

In this context, Dağtan's (2012) study draws the attention. In his study, he examined the washback effects of UDS by the data from 161 academicians through questionnaires which include open-ended, multiple-choice and 5 Likert-type items. The quantitative data were analysed through SPSS 17.0 Statistical Software, the qualitative data were analysed via content analysis. The findings showed that vocabulary questions were regarded as the most difficult ones, therefore, the candidates put more emphasis on improving the knowledge of vocabulary. While UDS enables the candidates to improve their reading comprehension skills, the other language skills are ignored. However, the participants scored the highest value to the item suggesting that in case of a change in the structure of the test, UDS should be redesigned in a way which will encompass four skills of the target language.

Another study in the same vein was carried out by Çakıldere (2013) who scrutinized the washback effects on KPDS and UDS tests on language learning of academic staff. The study conducted with 203 academic staff in a state university and the questionnaire and interviews were used as the main data gathering instruments. The participants were asked to fill in the questionnaires and then the interviews were made with the randomly chosen 4 academic staff. As result of the study, it was found that while KPDS and UDS had a positive effect on reading comprehension, the knowledge of grammar and vocabulary, the washback effect was on the negative side for listening, speaking, and writing skills. In addition, the researcher highlighted that even though such language skills as listening, speaking, and writing are not measured in these tests, they stated that they would have studied for them if the tests had involved them.

The negative washback effect of foreign language tests was also attested by Polat (2017) who investigated what kinds of effects the content of the test and the results have on the test-takers. To capture a detailed picture of the current condition, the mixed method research design was adopted in the study. The quantitative data collected from 2683 participants who are working as academic staff in 84 universities of the different regions and the qualitative data were obtained through semi-structured interviews. Such in Çakıldere (2013), the findings indicated that YDS is not considered as a practical foreign language test since it does not cover four skills of the target language, the difficulty of the test creates pressure on the participants and their language learning is negatively affected. The study attested that the content and the issues regarding the results (such as equivalence values of the test) should be reorganized.

2.10.2. Content and Changes in the Tests

Biltekin (2004) carried out a thesis study to find out the skills assessed in YDS. He paved the way for the study by asking the perceptions of the teachers of English working in the division of foreign language in high schools. In the light of the collected data, he analysed YDS examination administered in 1999. The findings showed that YDS mainly assessed reading comprehension skills of the test-takers. The main factor behind it was that it measured writing and speaking skills through reading since it was believed that it was really hard to interview them individually or ask them to write an essay and score them. In addition, the questions aiming to measure writing skills did not focus on active but passive knowledge of vocabulary. Moreover, he added that

even if they can, to some extent, measure writing performance, the internationally recognized TOEFL and the like tests utilized essay-writing tasks to measure it; by this way, they could get a clear image of the test-takers' writing performance.

Elaldı (2005), aiming at unearthing some difficulties and problems that the exam-takers face, examined the questions of KPDS. The study was carried out with the academicians working in a state university and the teachers offering proficiency test-oriented courses in the province of Sivas. The data proceeded in the study were collected through structured interviews, written questionnaires, and the questions of a selected KPDS test. The findings indicated that being proficient in reading and test strategies were the key factors to get desired scores in the test. According to the findings of the study, it was found that even if they are at the intermediate level of language proficiency, it does not guarantee them to succeed in the test, not because their poor knowledge of language skills in the target language but because they should be familiarized with the types of questions and should know the test strategies used in the test. Another finding of the study related to the time allocated for the test. The participants complained that they are forced to respond to all the questions at short notice, therefore they respond quickly and unconsciously just so as to reach to the end of the test on time. The interview conducted with the teachers uncovered that KPDS merely focuses on the knowledge of grammar and aims at assessing some specific language skills. Hence, the candidates studying for it are motivated to improve the knowledge of vocabulary and to have the knowledge of grammar.

Arslan (2005) also carried out a quite similar study in which she sought for practical solutions for the test-takers in UDS. Like Elaldı (2005), she collected the data through written questionnaires, structured interviews, and the questions of a selected UDS test. The written questionnaires were administered to 221 students and academicians and later on, the lecturers who prepare candidates for UDS were interviewed. The analysis of the collected data led her to the similar results with Elaldı (2005). According to the findings, it was observed that having strong knowledge of grammar and vocabulary, being proficient in reading comprehension, and being aware of the test techniques are the prerequisite factors which will enable candidates to succeed in the test. Besides, the researcher highlighted that reading is regarded as the fundamental practice to prepare for such kind of test.

Aşkaroğlu (2014) shed light on the differences in the contents of similar high-stakes tests. In his study, he compared such language tests as UDS (Examination of Inter-University Foreign Language) which was administered for the purpose of academic appointment and promotion in the fields of sciences, social sciences, and health sciences, KPDS (Examination of Foreign Language Proficiency of State Employees, which was also used for language allowance, and YDS (Examination of Foreign Language Proficiency) which was the combination of the previous two tests under a single title. He scrutinised the samples of those tests based on the level of difficulty, numeric values and the quality of the contents. The findings of the study showed that while all the tests consist of 80 multiple-choice items and the number of the questions for each section was nearly the same, the test-takers are asked to complete the previous two tests (UDS and KPDS) in 180 minutes, but YDS in 150 minutes. This finding led the research to the conclusion that the function of these tests changed from “proficiency” to “competition”. Based on the analysis of the sections, he also concluded that compared to the others, YDS consists of long and complex sentences and requires more knowledge of the target language.

The review of the available literature has shown that foreign language tests assessing only one skill of the target language have some consequences for learning and teaching processes. As stated above, they negatively affect the learning and practice of other skills. Since these kinds of tests are basically developed in a way which will cover reading comprehension and the knowledge of grammar and vocabulary, the productive skills (writing and speaking) and the receptive skill of listening are neglected. Therefore, test-takers become mechanical language learners that merely concentrate on the content of the test and practical test strategies.

Whereas previous studies were based on the data collected through questionnaires, structured / semi-structured interviews or oral/written forms and fundamentally investigate washback effects or contents of the tests, the present study focuses on actual performance of the participants in four skills of that language, explores participants’ performances in two kinds of tests, and examine the role of previous test (LYS-5 Foreign Language Test) in predicting participants’ scores in the second test (Test of English Language Proficiency). In this way, the data proceeded in the study are supposed to be purified from limitations based on self-reported data in which the problem may not fully be explored because of such issues as set response, low response

rate, lack of flexibility, misunderstanding of statements/questions, or social desirability bias. Along with the findings of previous research, such a predictive correlational research will contribute to the relevant field and have the characteristics of a report for the policymakers.



CHAPTER III

METHODOLOGY

3.1. Introduction

In this chapter, the methodology of the research will be explained and based on the purpose of the study, the research design, study group, data gathering tools, data collection and data analysis procedures will be explained in detail.

3.2. Research Design

The purpose of this study is to delve into the place of LYS-5 English Test on predicting ELT students' success in four skills of the target language. Considering the objective of the study, the review of related literature on research methodologies has attested that the predictive correlational research was considered as a decent research when the overall objective of the study is to investigate the relationship among the variables and to find out a regression equation that helps make predictions about a population in the existence of a relationship (Cohen, Manion, & Morrison, 2007).

3.3. Study Group

The study group consists of 81 (Female=56, Male=25) undergraduate ELT students who are either newly-enrolled or studying in the Department of English Language Teaching during the fall semester of the 2016-2017 academic year at a state university located in a large province in Southern Turkey. Of these students, 25 students enrolled in 2015 and had training either in the preparatory school (23) or faculty (2), 56 students enrolled in 2016 and did not have any training in the university but took the proficiency test administered by the university. Based on the proficiency test results, it was determined that while 22 students will attend the extensive language training in the preparatory school, 34 students will directly go for training in the faculty. All of the participants volunteered to take part in the study. Their ages range from 17 to 27.

3.4. Data Gathering Instruments

Two data gathering instruments were used to reach decent data in the study. A proficiency test entitled “Test of English Language Proficiency (TELP)” (see Appendix A) developed by the researcher was administered to the study group and the participants’ scores of LYS-5 English Test were utilized to predict the effect of this test on students’ success in the TELP.

3.4.1. Proficiency Test (Test of English Language Proficiency)

As the main data collection tool, the proficiency test entitled “Test of English Language Proficiency (TELP)” was used to assess students’ proficiency in English. The content of the test is as follows:

3.4.1.1. Developing and Piloting Procedures of TELP

Prior to starting to develop the test, the objectives and context of the test were determined. Since it is supposed the students of the study group to be at least B2 level to maintain their training in the faculties, the level of the test was specified as B2. The draft form of the test was shaped in line with the description in CEFR that characterized B2 (vantage) level students as follows:

“Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and Independent disadvantages of various options.”

(p.24)

Four internationally recognized English proficiency tests (TOEFL, IELTS, TOEIC, PTE) were inclusively studied by the researcher, a question pool was created and potential items were put in. Concerning the target skills and strategies to be used, the items were carefully prepared. Upon developing the final draft, it was sent for experts’ review who consisted of four academicians (three in the department of English Language Teaching, one in the department of measurement and evaluation). After the review, the content validity evidence was checked by the experts. All the experts acknowledged that the test has sufficient content validity evidence. Later, it was piloted to 28 students who were also studying in the Department of English Language

Teaching. Since it consisted of different parts and each part assessed a unique skill, some reliability and validity evidence were put forward.

The first two parts of test consisted of reading and listening questions. While the questions in these parts had varying question types, they were all binary (1-0) items. Taking the studies carried out with similar tests in the literature into consideration, it was revealed that item analysis was a prerequisite for these questions.

The item analysis is a procedure which contributes to the fairness of the test and analyses the factors such as item difficulty, item discrimination, standard deviation, KR-21 values and other descriptive data for each and every item (Professional Testing Inc., 2006).

Item difficulty refers to the percentage of examinees that answered the question correctly (Boopathiraj & Chellamani, 2013). The values range from 0 to 1 and the higher value is, the easier question will be. While the items with .90 p values are categorised as very easy items, the items with below .20 are considered as very difficult in the literature. For this reason, .50 p-value is accepted as optimum level. (Boopathiraj & Chellamani, 2013).

Item discrimination, in other respects, measures the difference between the percentage of the upper (correctly responded) and lower (responded incorrectly) groups (Sim & Rasiyah, 2006). In other words, it refers to the power of discrimination to the degree which the items successfully discriminate participants regarding the ability assessed by the item (Boopathiraj & Chellamani, 2013). Kuder Richardson -21 (KR 21) is used to measure the internal consistency of test scores obtained concurrently (Büyüköztürk, 2016) and commonly preferred when the tests consist of binary items. Considering all these factors, the data obtained from reading and listening parts were analysed through psychometric data analysis software.

According to the findings, item difficulty, item discrimination and Kuder-Richardson 21 (KR-21) scores of the pilot test were checked and the items which were away from the optimum levels were eliminated. The final questions' item difficulty (p), item discrimination (D), Kuder-Richardson 21 (KR-21) scores were presented in Table 3.1. The data collected from writing and speaking parts were tested by two raters.

Table 3.1 Test of English Language Proficiency Questions Item Analysis Result

Item	p	D
R1	0.4483	0.6833
R2	0.5172	0.6307
R3	0.4138	0.6424
R4	0.3793	0.6039
R5	0.3793	0.5671
R6	0.5517	0.7885
R7	0.4138	0.7026
R8	0.3448	0.6953
R9	0.3448	0.5813
R10	0.4138	0.5193
R11	0.6552	0.5958
R12	0.3103	0.3269
R13	0.5862	0.5434
R14	0.5862	0.7347
R15	0.4828	0.7138
R16	0.4138	0.7172
R17	0.4483	0.6469
R18	0.3448	0.7709
R19	0.6207	0.6465
L1	0.4828	0.6314
L2	0.3448	0.6932
L3	0.3448	0.6080
L4	0.5517	0.7833
L5	0.4483	0.7636
L6	0.5862	0.4381
L7	0.3793	0.6065
L8	0.4138	0.6058
L9	0.4138	0.5461
L10	0.6207	0.5136
L11	0.7586	0.7233
L12	0.4138	0.4865
L13	0.4483	0.5195
L14	0.6897	0.4610

p: Item Difficulty

D: Item Discrimination

KR21 (Reading):0.7819 / Listening: 0.7207

The raters as crucial components played a significant role in the evaluation process. Interrater reliability describes the concern that the participants' scores may differ from one rater to another (Leydens & Thompson, 1997; Moskal, 2000). Regarding the scores of two raters, inter-rater reliability was tested. Since writing and speaking parts were adapted from PTE and TOEIC Tests, the raters used scoring rubrics, which were developed by either PTE or TOEIC experts to assess students' products.

3.4.1.2. Content of TELP

There are four sections in the test. It was administered in two separate sessions on the same day: Morning Session, Afternoon Session (see Table 3.2.). In the morning session, examinees were asked to answer questions in the skills of reading, listening, and writing; in the afternoon session, they completed the speaking section.

Table 3.2. Description of TELP Test

Test of English Language Proficiency	
Morning Session	
Skills:	Reading, Listening, Writing
Time:	~90 minutes
Afternoon Session	
Skills:	Speaking
Time:	~10 minutes

3.4.1.2.1. Reading Section

This section of the test covers two reading tasks. Texts were adapted from IELTS academic and TOEFL practice tests, and new questions were prepared for each task. Extensive analysis of reading passages of internationally recognized language tests (TOEFL, TOEIC, IELTS, PTE) showed that texts with a length of 500-800 words are commonly used to assess examinees' reading comprehension skills. Taking this fact into consideration, special attention was paid to the length of the texts.

Task 1 (Careful reading)

In this part, examinees were asked to read the given passage and answer the questions in 20 minutes. They included different aspects of reading skills such as reading for gist, reading in detail, grasping specific points, matching information and passage-completion.

Task 2 (Careful reading)

Examinees were required to read a text and answer 13 questions for two separate tasks. The text was divided into seven paragraphs and examinees were asked to match headings for each paragraph in the first task. Later, they were required to match information or features and write TRUE/FALSE or NOT GIVEN for every statement.

Table 3.3. Description of Reading Comprehension Section

N of tasks:	2
Length of texts	702/589 (words)
Time allocated:	40 minutes
Types and number of questions:	Multiple-choice: 6 Matching heading:7 True/False/Not Given:7
Reading Skills & Strategies:	Scan and skim a text Understand specific points Match specific pieces of information /features

3.4.1.2.2. Listening Section

There were two tasks in the section. The audios were selected in line with the nature of IELTS Academic and questions were prepared based on the transcript and flow of audios. The examinees heard each part once and answered 14 questions. For each task of the section, there was time for them to look through the questions and time to check their answers.

Task 1 (Conversation)

The examinees listened to a conversation in which a woman was calling an insurance company to report an accident. They began in the first part of the conversation to answer the questions 1-7 and had some time to look at the questions 8 and 9. In the end, they were given half a minute to check their answers. Since the answers of the first part might contain incomplete answers or spelling mistakes, they were scored by two experienced raters. The second part was coded as binary items “1” and “0” and graded by one rater.

Task 2 (Lecture)

In this task, the examinees listened to a lecture talking about when-which elephants communicate. First, they had some time to look at the questions and they answered as they listened because they would not hear the recording a second time. The questions were in the form of multiple-choice; therefore, their answers were graded only by one rater.

Table 3.4. Description of Listening Comprehension Section

N of tasks:	2
Length of recordings	Conversation: 7 minutes 30 seconds Lecture: 7 minutes 39 seconds
Time allocated:	~ 20 minutes
Types and number of questions:	Form completion: 7 Short-answer questions:2 Multiple-choice :5
Listening comprehension skills & Strategies	Understand the topic and instructions Follow lines of argument Distinguish between fact and opinion Understand main reasons for and against an idea Recognize emphasis through intonation and stress

3.4.1.2.3. Writing section

There were two tasks in the section. They were adapted from the format of Pearson Test of English (PTE) Academic Writing Test.

Task 1 (Summarizing a written text)

The examinees needed to write a summary of a given text in one sentence. They had 10 minutes to complete this task. They were asked to ensure that the summary had better cover the main points of the text in a full, single sentence of no more than 25 words. Their answers were scored by two experienced raters who utilized PTE scoring rubric for academic writing (see Appendix G).

Task 2 (Writing an essay)

In this task, the examinees were asked to write an opinion essay for a topic. They had 20 minutes to plan, write and revise their essay of 200-300 words. Their responses were judged on how well they developed a position, organized their ideas, presented supporting details, and controlled the elements of standard written English. PTE academic writing rubric (see Appendix J) was used by two experienced raters to assess examinees' responses.

Table 3.5. Description of Writing Section

N of tasks:	2
Time allocated:	30 minutes
Types and number of questions:	Summarizing written text Writing essay
Writing skills & Strategies	Identify a writer's purpose, tone or attitude Comprehend explicit and implicit information Synthesize information Write a summary Organise sentences and paragraphs in a logical way Support an opinion with details

3.4.1.2.4. Speaking Section

This section of the test encompassed six different tasks (reading a text aloud, describing a picture, responding to questions (3), and expressing an opinion). They were all adapted from TOEIC tests. Concerning reliability and validity of speaking, they were not practised face-to-face, but rather in one-way talks. Taking previous experience of the researcher with TOEFL iBT into account, the exam rooms' echo was initially tested with a different number of students before piloting and it was found acceptable to perform the tasks in groups of 10 since the examinees became distracted in larger groups because of the echo in the exam rooms. Under the supervision of nine proctors, the sessions were simultaneously held in exam halls. Since there was no dual conversation between the proctors and examinees but they acted as a computer, there was no need to find English-speaking proctors. Regarding the inadequate infrastructure of laboratories, the examinees were asked to record their voices via mobile phones and send them to the wireless platform before they left the exam rooms. The tasks were separately handed to the examinees (e.g. they took the second task after completing the first one.)

Task 1 (Reading a text aloud)

In this task, the examinees were asked to read a given text aloud. In this way, their pronunciation, intonation and stress were assessed through two TOEIC scoring rubrics (See Appendix B and C) by two experienced raters. They had 15 seconds for preparation and 45 seconds to read the text aloud.

Task 2 (Describing a picture)

In this task, a black and white picture was handed in the examinees and they were asked to prepare their talks in 30 seconds and speak about the picture for 45 seconds. Specially developed TOEIC scoring rubric (see Appendix D) was used as a criterion and two experienced raters scored the examinees' responses.

Tasks 3-4-5 (Responding to questions)

Compared to other tasks, it included three interconnected sub-tasks. While no preparation time was provided, they had 30 seconds (15 seconds for per task) for the

first two questions and another 30 seconds to respond to the last question. The records were scored by two experienced raters by using the relevant TOEIC scoring rubric (see Appendix E).

Task 6 (Expressing an opinion)

In this task, the examinees were asked to give their opinion on a specific topic. They had 15 seconds to prepare and 60 seconds to speak about the topic as much as they could. The records were scored by two experienced raters by benefitting from the relevant TOEIC scoring rubric (see Appendix F).

3.6. Description of Speaking Section

N of tasks:	6
Time allocated:	~ 4.5 minutes
Types and number of questions:	Read a text aloud: 1 Describe a picture: 1 Respond to questions:3 Express an opinion:1
Speaking Skills & Strategies:	Pronunciation, Intonation and Stress Grammar, vocabulary, cohesion Relevance and completeness of the content

3.4.1.3. Raters

They were two teachers of English who had at least five years of teaching and testing in the target language. Before the piloting process, they were informed about the scoring rubrics used in the research. In total, they used seven different scoring rubrics for writing and speaking sections. During the piloting, Cohen's Kappa Analysis, the measure of inter-rater agreement when there are two or more raters (McHugh,2012), was computed to check whether such raters agree for each task, the results were presented in Table 3.7.

Table 3.7. Inter-rater reliability results

Tasks	Kappa Measure of Agreement
Summarize a text	.835
Write an essay	.85
Read a text aloud	Pronunciation .886
	Intonation and Stress .864
Describe a picture	.905
Respond to questions (1)	.867
Respond to questions (2)	.94
Respond to questions (3)	.91
Express an opinion	.915

Based on the suggestion of Cohen (1960), the values ranging from .80 to .90 can be interpreted as an indication of “strong agreement” and if they are above .90, that shows “almost perfect” agreement between the raters.

3.4.2. LYS-5 English Test Scores

LYS-5, one of the five higher education admission tests, is held in several languages such English, German or French and the candidates are required to complete it in 120 minutes (Council of Higher Education, 2016). It is organized and administered by OSYM (Student Selection and Placement Centre) which is an institute in charge of preparing and administering these tests. While placing ELT students into universities, scores of AOBP (High School Achievement Grade Point), scores of YGS (University Entrance Examination) and LYS-5 English Test are calculated to determine their overall placement scores but LYS-5 embodies the greatest effect with a percentage of % 65 among them (Council of Higher Education, 2016). LYS-5 English Test scores also constitute a momentous part of the study. Therefore, the scores of the participants were gathered on an individual basis under the supervision of the researcher. In order to both avoid the effect of YGS and AOBP scores and explore the pure effect of LYS-5 English Test in predicting students’ performance in four language skills, the present

study merely focused on LYS-5 English Test scores. Since there was no penalty for answering a question incorrectly in any of nationally or internationally recognised language tests (e.g. YDS, YOKDIL, TOEFL iBT, IELTS, PTE, and TOEIC), it was turned a blind eye to the number of incorrect responses in LYS-5 English Test, so the number of correct responses was multiplied by 1.25 to acquire a total score of 100 (see Appendix K). Upon gathering LYS-5 English Test scores, missing values and outlier's analysis were computed and they were replaced with the mean of series.

3.5. Data Collection Procedure

The data collection process started after receiving the permissions of the university to administer the TEFL test. Following the formal procedure, the test consisting of reading, listening and writing parts was administered to the pilot group concurrently. The pilot group comprised of 28 (13=Female, 15=Male) 2nd grade undergraduate ELT students. The reading part consisted of two tasks and the examinees had 40 minutes to complete them. In the listening part, there were two tasks which the examinees heard only once and took 20 minutes in total. The writing consisted of two tasks and the examinees had 30 minutes to complete them. In total, the first session took 80 minutes. Before applying the speaking test, the classroom conditions such noise and echo were controlled. As a result of noise and echo controls, it was found acceptable that 10 participants for each exam room were optimum for the speaking session. Therefore, the students were placed into different classrooms in the groups of 10. There were six tasks in the section and the examinees were asked to record their voices according to the objectives of the tasks. It took approximately 10 minutes to finish all the tasks. After that, they sent their voice records to the supervisor through hotspot provided by the university. Based on the collected data, the item analysis was performed for the binary items (1-0). Since writing and speaking parts were not categorised as binary, and these parts were scored by two experienced raters by using relevant rubrics, inter-rater reliability analysis was computed. Besides, experts' review was asked for the validity consideration. After analysing the data, a question (question 10) in the listening part which did not have optimum values in the item analysis, was removed and the test took its final form (see Appendix A).

In the second phase, the TEFL test was administered to the study group. Like in piloting, it was held in two sessions and the same procedures were followed. In the

final phase, the researcher collected LYS-5 English scores of the groups. Apart from the ones who had a different test called YOS (International Student Exam), the researcher reached the test results of all the group.

3.6. Data Analysis

The collected data were analyzed through SPSS 23.0 Quantitative Data Analysis Software. The correct responses were accepted as “1”, incorrect or missing responses were accepted as “0”. Since the ultimate purpose of the present study is to find out the role of LYS-5 Foreign Language Test in predicting students’ performances in four skills of the target language, the regression analysis was performed to calculate values of each variable. Based on the assumptions of regression analysis, the normality distribution values of the variables were analyzed. After the simple linear regression analysis was computed to estimate prediction values of LYS-5 Foreign Language Test for the total scores of TELP test, the multiple regression analysis was performed in order to calculate prediction values for every skill. In the second phase, TELP test scores of the participants were analyzed by splitting their demographic information as “attended or exempt from the preparatory school”. The analysis results were presented as mean values in tables.

Table 3 .8. Descriptive Statistics for Different Variables

		TELP	LYS-5	Writing	Reading	Listening	Speaking
N	Valid	81	81	81	81	81	81
	Missing	0	0	0	0	0	0
Mean		52.974	83.090	14.891	12.313	11.848	18.125
Median		51.995	83.026	14.913	11.842	12.500	18.210
Mode		21.475	77.500	17.185	10.526	10.714	19.792
Std. Deviation		13.950	5.659	5.171	4.089	3.980	3.181
Variance		194.615	32.021	26.742	16.720	15.839	10.121
Skewness		-0.013	-0.071	-0.284	-0.031	-0.179	-0.167
Kurtosis		-0.554	0.320	0.118	-0.834	-0.373	-0.188

Table 3.8 presents the findings of the normality distribution values of the collected data. As can be understood from the values of the central tendency measures (the coefficients of skewness and kurtosis), the values for all the variables are close to “1” and lie within the limits of ± 1 . As Büyüköztürk, Çokluk and Köklü (2013) point out, the values lying within the acceptable limits of ± 1 can be regarded as an implication of normal distribution for the variables.



CHAPTER IV

FINDINGS

4.1. Introduction

In this chapter, the analysis of the data obtained in the research process is followed by findings and interpretations of these findings.

4.2. Findings on the Values of LYS-5 Foreign Language Test in Predicting Four Language Skills

In this part, the students' scores in LYS-5 Foreign Language Test and TELP (Test of English Language Proficiency) were analysed through simple linear and multiple regression analysis respectively and the findings were interpreted in line with the results. Since the ultimate goal of the present study is to examine the pure effect of LYS-5 English Test in predicting students' success in four language skills of the target language, the students who enrolled in 2015 and had training either in the preparatory school or in the faculty were excluded from this analysis. Therefore, the regression analyses were merely computed with the students of 2016. While it was aimed at reaching prediction values based on the total scores of TELP and LYS-5 English Test by the simple linear regression analysis, the values for each separate skill were acquired via multiple regression analysis.

Table 4.1 Model Summary and ANOVA Results

	Sum of					
	Squares	df	R	R ²	F	p
Regression	780.379	1	.281	0.079	4.614	0.036
Residual	9133.479	54				
Total	9913.858	55				

Table 4.1 illustrates the relationship between the dependent (TELP) and independent (LYS-5 English Test) variables. As presented in Table 2, the relationship was estimated as 0.281 and it is an indicator of a low-level relationship. The participants' scores in LYS-5 English Test explain the total variance of TELP (Test of English

Proficiency) at the rate of 7.9%. Depending on the findings, H₁ hypothesis (LYS-5 English Test scores are correlated with TELP scores) is accepted and the model is found “significant” $F(1, 54) = 4.614, p < 0.05$.

Table 4.2 Regression Model

Model		B	Std. Error	β	t	p
1	(Constant)	-20.053	33.488		-.599	0.552
	LYS-5 Score	0.842	0.392	0.281	2.148	0.036

Considering the significance value (p) of the regression coefficient, LY5 English Test is found as a predictor of success in TELP, $p > 0.05$. However, when R (.281) and R² (7 %) values are taken into consideration, it may not be possible to regard LYS-5 English as a strong predictor even if the equation is significant.

Table 4.3 Multiple Relations CI Values

Model	Dimension	Eigenvalue	Variance Ratios					
			Condition Index	Constant	Writing	Reading	Listening	Speaking
1	1	4.753	1.000	0.00	0.00	0.00	0.00	0.00
	2	0.116	6.406	0.00	0.35	0.04	0.45	0.00
	3	0.071	8.178	0.00	0.52	0.64	0.10	0.00
	4	0.045	10.263	0.20	0.12	0.32	0.43	0.13
	5	0.015	17.587	0.79	0.00	0.00	0.01	0.87

When the cases have higher than 30 CI value, it is interpreted that the variables have multiple relationships among themselves. As can be understood from Table 6, all of the variables have CI values smaller than 30. Hence, it attests that there is no multiple collinearity condition in the given dataset. This finding contributes and fulfils one of the assumptions to compute regression analysis.

Table 4.4 Model Summary and ANOVA Results

	Sum of Squares	df	R	R ²	F	p
Regression	209.966	4	.437	0.191	3.009	0.026
Residual	889.712	51				
Total	1099.678	55				

The relationship between the dependent and independent variable was calculated as 0.437. This ratio is the indicator of a moderate-level relationship. LYS-5 English Test scores of the participants predict their scores in reading, writing, listening and speaking sections of TELP at the rate of 19 %. In the light of the findings, H₁ hypothesis (LYS-5 English Test scores has an effect of the TELP skills' scores) is accepted and the developed model is considered as significant, $F(4, 51) = 3.009, p < 0.05$.

Table 4.5 Regression Model

Variable	B	Std. Error	β	t	p	Zero-order	Partial	Tolerance	VIF
(Constant)	76.834	3.387		22.687	0.000				
Writing	-0.016	0.113	-0.019	-0.139	0.890	0.125	-0.020	0.861	1.161
Reading	0.445	0.160	0.379	2.785	0.008	0.407	0.363	0.858	1.166
Listening	0.134	0.149	0.120	0.899	0.373	0.182	0.125	0.890	1.124
Speaking	0.112	0.202	0.078	0.552	0.584	0.222	0.077	0.789	1.267

As can be seen in Table 4.5, there is a very low positive correlation between writing scores and LYS-5 English Test scores ($R=0.125$); moderate-level positive correlation between reading scores and LYS-5 English Test scores ($R=0.407$); and low positive correlation between listening and speaking scores, and LYS-5 English Test scores ($R=0.182$; $R=0.222$, respectively). When the points of other variables are controlled, a considerable change does not occur in the ratio of the correlation.

Based on the t test results regarding the significance of the regression coefficient, apart from reading, other sub-dimensions (writing, listening, and speaking) are not a strong predictor of LYS-5 English Test.

The regression equation emerged as a result of the analysis for predicting LY5 English Test as follows:

$$\text{LYS-5 English Test} = 76.834 - 0.016(\text{Writing}) + 0.445(\text{Reading}) + 0.134(\text{Listening}) - 0.112(\text{Speaking})$$

4.3. Findings on Reading Comprehension Scores of the Participants

In this part, the scores of the participants in the reading comprehension section were analysed and the results were presented and interpreted by means of mean values, and

standard deviation. Unlike the regression analysis, the findings presented below cover the whole group (both 2015 and 2016). The results are firstly presented for each particular year and total values of the whole group are eventually shown.

Table 4.6 Descriptive Statistics for Reading Section in TELP

Test/Skill	Year	Preparatory/Proficiency Test	\bar{X}	SD	N
Reading	2015	Attended P.S. *	13.902	4.100	23
		Exempt from P.S. **	10.526	9.304	2
		Total	13.632	4.460	25
	2016	Exempt from P.S.	12.190	3.582	34
		Not exempt from P.S.***	11.005	4.112	22
		Total	11.725	3.808	56
	Total	Attended P.S.	13.902	4.100	23
		Exempt from P.S.	12.098	3.837	36
		Not exempt from P.S.	11.005	4.112	22
		Total	12.313	4.089	81

* Attended Preparatory School

**Exempt from the Preparatory School

*** Not exempt from Preparatory School

Reading scores of the participants are presented in Table 4.6. As can be seen in the table, the participants attending the P.S. are more successful than the ones who are exempt (\bar{x} = 13.902, \bar{x} = 10.526, respectively). Even though there is not any participant attending the P.S. In 2016, the ones who are exempt from the P.S. thanks to their success in the proficiency test outperformed compared to the second group (\bar{x} = 12.190, \bar{x} = 11.005). When the findings are interpreted in total, it is noticeable that while the participants going through training in the P.S. have the greatest average, the ones who are not exempt from the preparatory school displayed the lowest average among the groups.

4.4. Findings on Listening Comprehension Scores of the Participants

In this part, the scores of the participants in the listening comprehension section were analysed and the results were presented and portrayed by means of mean values, standard deviation and number of participants in the years of 2015 and 2016. The results are initially illustrated for each particular year and total values of the whole group are eventually shown.

Table 4.7. Descriptive Statistics for Listening Section in TELP

Test/Skill	Year	Preparatory/Proficiency Test	\bar{X}	SD	N
Listening	2015	Attended P.S. *	13.231	3.794	23
		Exempt from P.S. **	11.607	1.263	2
		Total	13.101	3.669	25
	2016	Exempt from P.S.	12.027	4.154	34
		Not exempt from P.S. ***	10.146	3.588	22
		Total	11.288	4.016	56
	Total	Attended P.S.	13.231	3.794	23
		Exempt from P.S.	12.004	4.041	36
		Not exempt from P.S.	10.146	3.588	22
		Total	11.848	3.980	81

* Attended Preparatory School

** Exempt from the Preparatory School

*** Not exempt from Preparatory School

When the listening section averages of the participants in Table 4.7 are examined, it is possible to see that the condition of attending the P.S. showed a significant role in 2015. In this year, while the listening averages of the first group (attending P.S.) was measured as 13.231, the average in the second group was calculated 11.607. While the current situation in this manner in 2015, the results of the participants in the following year revealed that listening averages are correlated with the condition of exempt or not exempt from the preparatory school. Compared to the ones who are not exempt from the preparatory school (\bar{x} = 10.146), the students who are exempt from it reached a higher average in the listening section (\bar{x} = 12.027). Nevertheless, their averages were lower than the participants who attended the P.S. in 2015 (\bar{x} = 13.231).

4.5. Findings on Writing Scores of the Participants

In this part, the scores of the participants in the writing section were analysed and the results were presented and depicted by means of mean values, standard deviation and number of participants in the years of 2015 and 2016. The results are firstly presented for each particular year and total values of the whole group are eventually shown.

Table 4.8. Descriptive Statistics for Writing Section in TELP

Test/Skill	Year	Preparatory/Proficiency Test	\bar{X}	SD	N
Writing	2015	Attended P.S. *	16.074	4.649	23
		Exempt from P.S. **	13.776	4.821	2
		Total	15.890	4.603	25
	2016	Exempt from P.S.	14.528	5.779	34
		Not exempt from P.S. ***	14.317	4.842	22
		Total	14.446	5.385	56
		Attended P.S.	16.074	4.649	23
	Total	Exempt from P.S.	14.487	5.673	36
		Not exempt from P.S.	14.317	4.842	22
		Total	14.891	5.171	81

* Attended Preparatory School

** Exempt from the Preparatory School

*** Not exempt from Preparatory School

Table 4.8 marks descriptive values of the participants in writing section of TELP. It is clear that the participants who enrolled in the university and attended the P.S. training in 2015 scored higher (\bar{x} = 16.074) in writing section than the ones who passed the proficiency test administered by the school and were exempt from the preparatory school within the same year (\bar{x} = 13.776). In 2016, since the sampling merely consists of the participants who are either exempt or not exempt from the preparatory school, the results are interpreted based on this fact. It can be understood that there is a slight difference between the participants (\bar{x} = 14.528, \bar{x} = 14.317, respectively) When the scores of the participants in the years 2015 and 2016 are compared, the participants that attended the P.S. achieved the highest average among them. It is followed by the participants who exempt from the preparatory school in 2016, and who are not exempt from the preparatory school in 2016, and lastly the ones who are exempt from the preparatory school in 2015.

4.6. Findings on Speaking Scores of the Participants

In this part, the scores of the participants in speaking section were analysed and the results were presented and interpreted by means of mean values, standard deviation

and number of participants in the years of 2015 and 2016. The results are firstly presented for each particular year and total values of the whole group are eventually shown.

Table 4.9. Descriptive Statistics for Speaking Section in TELP

Test/Skill	Year	Preparatory/Proficiency Test	\bar{X}	SD	N
Speaking	2015	Attended P.S. *	19.792	2.590	23
		Exempt from P.S. **	19.271	5.156	2
		Total	19.750	2.698	25
	2016	Exempt from P.S.	18.210	2.802	34
		Not exempt from P.S.***	16.146	3.263	22
		Total	17.399	3.132	56
	Total	Attended P.S.	19.792	2.590	23
		Exempt from P.S.	18.269	2.867	36
		Not exempt from P.S.	16.146	3.263	22
		Total	18.125	3.181	81

* Attended Preparatory School

**Exempt from the Preparatory School

*** Not exempt from Preparatory School

Unlike the other skills, there is merely a slight difference between the averages of the participants who attended or were exempt from the P.S. in 2015 (\bar{x} = 19.792; \bar{x} = 19.271; respectively). However, in 2016 there is a gap between two groups. In contrast with the ones who are exempt, the participants who passed it and were exempt from the P.S. demonstrated better performance in speaking section (\bar{x} = 16.146; \bar{x} = 18.210, in return). Yet, they could not attain the average of the ones attending the P.S. in the previous year (\bar{x} = 19.792).

4.7. Findings on Overall Test Scores of the Participants

In this part, the total scores of the participants in the Test of English Language Proficiency (TELP) were analysed and the results were presented and depicted by means of mean values, standard deviation and number of participants in the years of 2015 and 2016. The results are firstly presented for each particular year and total values of the whole group are eventually shown.

Table 4.10. Descriptive Statistics for TELP Scores

Test/Skill	Year	Preparatory/Proficiency Test	\bar{X}	SD	N
TELP	2015	Attended P.S. *	55.687	15.184	23
		Exempt from P.S. **	55.181	18.019	2
		Total	55.647	14.996	25
	2016	Exempt from P.S.	53.452	13.738	34
		Not exempt from P.S.***	49.200	12.807	22
		Total	51.781	13.426	56
	Total	Attended P.S.	55.687	15.184	23
		Exempt from P.S.	53.548	13.689	36
		Not exempt from P.S.	49.200	12.807	22
		Total	52.974	13.950	81

* Attended Preparatory School

** Exempt from the Preparatory School

*** Not exempt from Preparatory School

Table 4.10 depicts the picture of the participants' overall performances in TELP. As it is shown, there is a minor difference between the averages of two groups in 2015. While the participants who initially failed the proficiency test and attended the preparatory school displayed a performance at the rate of 55.687; the ones who were exempt from it and started their training in the faculty reached an average "55.181". For the new coming participants, the current situation was different from the findings of 2015. In this year, there is a considerable difference in the averages of the groups. As in other skills, the participants who were exempt from the preparatory school had less average than the other (\bar{x} =49.200; \bar{x} =53.452; in return). Nevertheless, their average was lower than both groups of 2015.

CHAPTER V

CONCLUSION

5.1. Introduction

The present study aimed to unearth the role of LYS-5 English Test in predicting the students' success in four skills of the target language and to depict the picture of the students' current readiness for undergraduate education. To attain this goal, a correlational survey research was adopted to a study group involving 82 undergraduate students who were either newly-enrolled or studying at the Department of Foreign Language Education in the Faculty of Education in a state university of a large province in southern Turkey. Two types of data gathering instruments were used in order to collect data: Test of English Language Proficiency and LYS-5 English Test scores. As for the analysis of the collected data, the statistical data analysis techniques were utilized, the findings were interpreted and thus answers concerning the objective of the research were sought.

In this chapter, the results obtained from the analysis of the data gathered through data gathering instruments are discussed and interpreted with reference to the relevant literature, recommendations are put forward for practitioners, policymakers, future researches and researchers.

5.2. Results and Discussion

The research findings revealed that although the students had considerably high scores in LYS-5 English Test, their scores did not show similarity with the scores of the TELP which is a four-skill assessment test. While their average score in LYS-5 English Test was 83,090, it was only 52,974 in TELP test. When the students' scores were analysed based on their current condition (exempt, not exempt, or attended the preparatory), the findings of the present study indicated that the ones who attended the preparatory school displayed higher scores on average and in four skills separately. This finding can be interpreted that attending the preparatory school is a distinctive feature in terms of the mastery of the target language. When the status of the students' proficiency tests is taken into account, it is clear that they do not correctly determine students' general

level of proficiency in the target language since the score averages of both groups in TELP test are nearly the same. The reason behind this finding can be pertinent to the content of the test which the students took. The proficiency test developed by the university where the present study was conducted did not measure speaking skills of the test-takers, therefore it might be difficult to predict students' overall performance in the target language by making inferences from the other skills.

To have a clear-cut picture of LYS-5 English Test on students' success in language skills, the students' scores were analysed regarding the four skills separately. The findings showed that LYS-5 English Test is a significant predictor of students' reading comprehension skills. Taking into account that LYS-5 English Test primarily consists of reading comprehension skill, and vocabulary and grammar, this finding of the research coincides with the nature of LYS-5 English Test itself. This finding of the present study also shows similarity with the previous studies carried out by Biltekin (2004), Elaldı (2005), Arslan (2005), Sevimli (2007), Yıldırım (2010), Dağtan (2012), Çakıldere (2013), Sayin and Aslan (2016) and Hatipoğlu (2016), who uncovered the positive effect of such language tests on reading comprehension skills of the participants.

As a result of the findings, LYS-5 English Test was not found as a strong predictor of students' writing performance, as expected. Since it does not measure or cover students' writing performance, this finding was regarded as a potential outcome of the analysis. When the standardized beta values of the multiple regression analysis are taken into account, it is obvious that LYS-5 has the lowest prediction value for writing skill. Even though it is believed that the students improve and have a wide range of the knowledge of vocabulary while preparing for LYS-5 English Test, their scores in TELP test point out that the knowledge of vocabulary itself is not sufficient to produce an intelligible writing task which requires not only knowledge of vocabulary but also knowledge of grammar and structure for the given task. This finding of the present study attests the results of Biltekin (2004) who pointed out that the productive skill of writing cannot fully be measured via a multiple-choice test which mainly consists of reading comprehension content and endeavours to measure test-takers' writing performance through reading. In addition, this finding has also correlated some researchers (Hatipoğlu,2016; Sayın & Aslan, 2016; Sevimli, 2007; Yıldırım, 2010) who considered writing skill as one of the neglected skills in LYS-5 English Test.

Like in other skills, the findings did not show a significant change in listening section, though. As a result of the analysis, it was found that LYS-5 English Test did not significantly predict the students' listening performance. As stated for writing skill, it was also a potential outcome of the analysis because the listening skills are not measured neither directly nor indirectly in LYS-5 English Test (Karabulut,2007).

The last skill examined was speaking. As expected, LYS-5 English Test was not a significant predictor of the students' speaking performance. Although some sections which ask the students to give the best response to the given situations or dialogues were considered to improve their speaking skills by some sides, the findings of this research unveiled that there was no sense of measurement regarding the speaking performance in LYS-5 English Test. This finding is supported by the previous researchers (Hatipoğlu, 2016; Karabulut,2007; Sayın &Aslan,2016) who pointed out that speaking skills of the test-takers are not measured in the framework of LYS-5 English Test.

The recent studies recursively emphasized that receptive skill of listening and productive skills of writing and speaking are ignored by the national language tests; therefore, the test-takers fully concentrate on the mechanical characteristics of the target language rather than fundamental skills (Biltekin, 2004; Çakıldere, 2013; Polat, 2017)

Considering the findings of the present study, it should be highlighted that the students' scores in LYS-5 English Test can mislead them about their mastery of the target language. They may not be aware of their weaknesses and strengthens regarding the target language (Hatipoğlu, 2016). The present study showed how well the students represent the skills of the target language.

As for attending the preparatory school, the findings pointed out that compared to being exempt from the preparatory school, the ones who underwent an extensive language training program in the preparatory school showed higher performance in TELP test. It is mainly because they learn the target language in a communicative learning atmosphere and have a chance to practice their learning during the program. Moreover, their learning is measured by the final exams which encompass four main skills of that language; therefore, it creates a sense of motivation and positive washback effect for learners, though. However, it may not be valid for LYS-5 English Test since it is a form-based multiple-choice test and measures only some aspects of the target language.

5.3. Suggestions for Practitioners and Policymakers

The results of the study reveal that it is a distinctive point whether or not the students attended the preparatory school, so the proficiency tests administered by the universities play a crucial role in higher education. However, the differences in the proficiency tests across the universities draw our attention. While some universities in Turkey administer proficiency tests which cover four skills of the target language, others still insist on conducting a multiple-choice proficiency test which measures students' knowledge of grammar, vocabulary and reading comprehension skills. Since these tests are prerequisite to diagnose the students' weakness and strengths in the target language and determine the students' readiness for the faculties, the contents of the proficiency tests should be either redesigned or changed completely, or a central proficiency test, which covers four skills and is recognised by all the universities in Turkey, can be developed. In this manner, the students can be placed into the appropriate level and the ones who have adequately proficient in the target language can go through their training in the faculties and the others may undergo an extensive language training and start their training in their departments in the following year.

Considering the types of English tests, LYS-5 English Test functions as a placement test. However, it is limited to reading comprehension as a main skill, vocabulary and grammar knowledge as subskills. When we take the role of the test into account, it is clear that future pre-service teachers of English are determined through it. Therefore, an alternative type of test which assesses four skills of the target language should be developed in order to depict a detailed picture of students' overall performance in the target language and place them into the appropriate departments and divisions. Considering the difficulty of rater reliability in writing and speaking skills, the transformation can be started from receptive (reading and listening) to productive skills (writing and speaking). In recent years, OSYM has started to administer the well-known national foreign language examination (YDS) electronically. It can be indicative of having sufficient infrastructure for such a change.

While current practices show that OSYM successfully administers the tests in terms of security, reliability, and practicality, based on the findings of the present study, it is suggested that OSYM should perform studies to improve the content of the test in order to create construct and content validity, though. In this way, other fundamental

principles of testing such validity, authenticity and washback can completely be fulfilled.

Since LYS-5- English Test serves as a single scale which measures students' language performances so as to enrol in a university, the students tend to study grammar, vocabulary, and reading comprehension to prepare for it. Hence, it is generally seen that test-oriented language education is practiced in most of the high schools, and the skills, which are not measured by LYS-5-English test, are generally ignored by the students. The students switch off their learning in those skills and fully concentrate on doing exercises relevant to the content of the test because it is vitally important for their undergraduate education and constitutes a big portion of their overall scores while placing in a university. It is recommended that a possible change in the structure of LYS-5- English Test will directly change negative washback effects of the test into positive, rather than a test-oriented education, a skill-based education will be fostered, and the students will give an equal emphasis on all skills of the target language.

5.4. Recommendation for Further Research and Researchers

The present study attests that LYS-5 English Test is a not a strong predictor of the students' language skills of listening, writing and speaking, and their future performances in these skills. However, since it is administered to the study group, it is highly suggested that a further research should be carried out with a larger sample to strengthen or add on the findings of the present study.

It is also suggested that a further study should be conducted in order to compare the results of the participants attending from different universities and regions.

In addition, the present study scrutinizes the relationship between English Admission Test (LYS-5) and Test of English Language Proficiency (TELP) by focusing participants' score in both tests. In further researches, the effect of participants demographic features (e.g. school type they attended (private or public), number of English-speaking family member, age, gender, duration of living abroad etc.) on their scores in the similar tests can also be examined.

Finally, since the results of the present study are mainly based on the quantitative data, they can be supported with qualitative data such as semi-structured interviews and written expressions.

REFERENCES

- Alderson, J.C. & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Alderson, J.C. (1996). "The testing of reading." In C. Nuttall (Ed.) *Teaching reading skills in a foreign language*, 212-228. Oxford: Heinemann.
- Alderson, J.C., & Hughes, A. (1981). *Issues in Language Testing*. London: British Council English Language and Literature Division.
- Alderson, J.C., Clapham, C., & Wall, D. (1995) *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Alıcı, S.S. (2004). *The Opinions of Yıldız Technical University, Foreign Languages Department English Preparatory School Students' on the Quality of Education They Are Provided with*. Unpublished Master's Thesis, Yıldız Technical University, İstanbul.
- Allal, L. (1988). Vers Vers un élargissement de la pédagogie de maîtrise. Les propositions de la pédagogie de maitrise, Delachaux et Niestlé Neuchatel, pp.86-126
- Anastasi, A. (1988). *Psychological Testing (6th Ed.)*. New York: Macmillan.
- Andrews, S. (2004). *Washback and Curriculum Innovation*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Anthony, J.N., & Susan, M.B. (2005). *Education Assessment of Students*. New Jersey: Person Education Ltd.
- Arslan, E. (2005). *Analysing UDS (English Proficiency Exam for Academicians) Exam Questions* (Unpublished master's thesis). Sosyal Bilimler Enstitüsü/Cumhuriyet Üniversitesi.
- Audibert, S. (1980), "En d'autres mots...l'évaluation des apprentissages !", *Mesure et évaluation en éducation*, Vol. 3, pp. 59-64.
- Aydogan, H., & Akbarov, A. A. (2014). The Four Basic Language Skills, Whole Language & Intergrated Skill Approach in Mainstream University Classrooms in Turkey. *Mediterranean Journal of Social Sciences*, 5(9), 672-680.
- Bachman, L. F.& Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessment and justifying their use in the real world*. Oxford: Oxford University Press.
- Bachman, L.F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of Applied Psychology*, 78, 111–118.
- Başar, E. (2001). *Genel öğretim yöntemleri*. Samsun: Kardeşler Ofset ve Matbaa.
- Başaran, İ. E. (1998). *Eğitim psikolojisi (5. Baskı)*. Ankara: Aydan Web Tesisleri.
- Biggs, J. B. (1995). Assumptions underlying new approaches to educational assessment. *Curriculum Forum*, 4 (2), 1-22.
- Biggs, J.B. & Tang, C. (2007). *Teaching for Quality Learning at University (3rd Ed.)*. Maidenhead: McGraw Hill Education & Open University Press.
- Biltekin, M. B. (2004). *Yabancı Dil Öğretim Yöntem ve Yaklaşımlarının YDS'nin Ölçtüğü Beceriler Açısından İncelenmesi* (Yayımlanmamış yüksek lisans tezi). Sosyal Bilimler Enstitüsü/Ankara Ünivertesii.
- Bloom, B. (1995). *İnsan nitelikleri ve okulda öğrenme (2. Baskı)*. (Çeviren: Durmuş Ali Özçelik). Ankara: Milli Eğitim Basımevi.
- Bloom, B.S., J.T. Hasting and G.F. Madaus (1971), *Handbook on Formative and Summative Evaluation of Student Learning*, McGraw- Hill Book Co, New York.
- Boopathiraj, C., & Chellamani, K. (2013). Analysis if the test items on difficulty level and discrimination coefficient in the test for research in education. *International Journal of Social Science & Interdisciplinary Research*, 2 (2), 189–192.
- Brock, R. L. (1998). *The Language Teacher*. The 64th TOEIC Seminar in Tokyo.
- Brown, D. H. (2004) *Language Assessment. Principles and Classroom Practices*. New Jersey: Pearson Education.
- Brown, H. D. & Abeywickrama, P.(2010. *Language assessment, principles and classroom practices (2nd Ed.)* . White Plains, NY: Pearson Education, Inc.
- Brown, H.D. (1994). *Principles of Language Learning and Teaching*. Englewood Cliffs, NJ: Prentice Hall.
- Büyüköztürk, Ş. (2007), *Sosyal Bilimler için Veri Analizi El Kitabı (22. Baskı.)*, Ankara: Pegem A. Yayıncılık.

- Cambridge University Press. (n.d.). Cambridge Dictionary | English Dictionary, Translations & Thesaurus. Retrieved May 5, 2018, from <https://dictionary.cambridge.org/>
- Campbell, J. P. (1990) 'Modeling the Performance Prediction Problem in Industrial and Organizational Psychology', in M. D. Dunnette and L. M. Hough (eds), *Handbook of Industrial and Organizational Psychology*. PaloAlto: Consulting Psychologists Press. Vol. 1: pp. 687-732.
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In Richards, J. C., & Schmidt, R. W. (Eds.), *Journal of Language and Communication*, 2-27. London: Longman.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Carroll, B.J. (1980): *Testing communicative performance*. London: Pergamon.
- Carroll, J. (1991) *Cognitive abilities in foreign language aptitude: Then and now*. Englewood Cliffs, NJ: Prentice Hall.
- Cheng, L. & Curtis, A. (2004). *Washback or backwash: a review of the impact of testing on teaching and learning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Chomsky N. (1986). *Knowledge of Language: Its Nature, Origin and Use*. New York, NY: Praeger.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: M.I.T. Press.
- Cohen J. A. (1960). Coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20,37-46.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education (6th ed.)*. New York: Routledge.
- Colarelli, S. M., Dean, R. A., & Konstans, C. (1987). Comparative effects of personal and situational influences on job outcomes of new professionals. *Journal of Applied Psychology*, 72, 558-566.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, U.K: Press Syndicate of the University of Cambridge.
- Council of Higher Education. (2016). *2016 Öğrenci Seçme ve Yerleştirme Sistemi (ÖSYS) Kılavuzu* (pp. 1-120, Rep.). Council of Higher Education.

- Council of Higher Education. (2017). *2017 Öğrenci Seçme ve Yerleştirme Sistemi (ÖSYS) Kılavuzu* (pp. 1-120, Rep.). Council of Higher Education.
- Çakıldere, B. (2013). *Washback Effects of High Stakes Exams, KPDS And UDS (YDS), On Language Learning Of Academic Personnel (Nevşehir Case Study)* (Unpublished master's thesis). Institute of Educational Sciences.
- Çelik, S. (2014). *Communicative language teaching*. Ankara, Turkey: Eğiten.
- Dağtan, E. (2012). *Üniversitelerarası Kurul Yabancı Dil Sınavı'nın Geri Etki (Washback) Açısından İncelenmesi* (Dicle Üniversitesi Örneği) (Yayımlanmamış yüksek lisans tezi). Eğitim Bilimleri Enstitüsü/Dicle Üniversitesi.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of Language Testing*. Cambridge: CUP/UCLES.
- Desheng, C. and Varghese, A. (2013). Testing and evaluation of language skills. *In IOSR Journal of Research*.
- Downing, J., Trackray, D.V. (1971) *Reading Readiness*. London: University of London.
- Educational Testing Service. (2009). *The official guide to the TOEFL iBT*. McGraw-Hill Education.
- Educational Testing Service. (2016). *Examinee Handbook Speaking & Writing (TOEIC)* [PDF]. Educational Testing Service.
- Elaldı, Ş. (2005). *Analysing KPDS (English Proficiency Exam for State Employees) Exam Questions* (Unpublished master's thesis). Sosyal Bilimler Enstitüsü/Cumhuriyet Üniversitesi.
- EPI. (2017). EF EPI 2017 RANKINGS (pp. 1-48, Rep.). Retrieved from <https://www.ef.com/~/media/centralefcom/epi/downloads/full-reports/v7/ef-epi-2017-english.pdf>.
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1-4.
- Fidan, N. (1986). *Okulda öğrenme ve öğretme*. Ankara: Kadioğlu Matbaası.
- Frederiksen, J. R., & Collins A. (1989). A system approach to educational testing. *Educational Researcher*, 18(9), 27-32.

- Garrett, T. (2008). Student-Centered and Teacher-Centered Classroom Management: A Case Study of Three Elementary Teachers. *Journal of Classroom Interaction*, 43(1), 34-47.
- Gilfert, S. (1996). A Review of TOEIC. *The Internet TESL Journal*, 2 (8). Retrieved August 26, 2002 from <http://iteslj.org/Articles/Gilfert-TOEIC.html>.
- Gronlund, N. E. (1998). *Assessment of student achievement (6th Ed.)*. Boston: Allyn & Bacon.
- Habaci, I., Kaya, I., Erdik, C., Adiguzelli, Y., & Sivri, M. (2011). Motivation Deficiency Experienced in Foreign Language Teaching in Turkey: Elimination Ways and Suggestions. *World Applied Sciences Journal*, 15 (3) (1818-4952), 400-406.
- Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behaviour and Human Performance*, 16, 250–279.
- Halliday, M.A.K. (1978). *Language as a Social Semiotic. The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Hamayan, E. (1995). Approaches to alternative assessment. *Annual Review of applied linguistics*. New York: Cambridge University Press. 212–226.
- Hamp-Lyons (1998). *Language testing and evaluation*. Retrieved from www.iosrjournals.org on 12/05/2018
- Harman, G., & Çelikler, D. (2012). A Review Study about Important of Readiness in Education. *Journal of Research in Education and Teaching*, 1(3), 147-156.
- Harmer, J. (2007). *The Practice of English language teaching*. Harlow: Pearson Education.
- Hatipoğlu, Ç. (2016). The Impact of the University Entrance Exam on EFL Education in Turkey: Pre-service English Language Teachers' Perspective. *Procedia - Social and Behavioral Sciences*, 232, 136-144
- Heaton, J. B. (1988). *Writing English Language Tests*. New York: Longman Inc.
- Heaton, J. B. (1990). *Classroom Testing*. London: Longman.
- Heaton, J.B. (1975). *Writing English Language Tests*. London: Longman.
- Wells, C. S., & Wollack, J. A. (2003). An Instructor's Guide to Understanding Test Reliability. Retrieved from <https://testing.wisc.edu/Reliability.pdf>.
- Hinkel, E. (2006). *Current Perspectives on Teaching the Four Skills*. TESOL Quarterly.

- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hughes, A. (1996). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hughes, A. (2003). *Testing for language teachers (2nd Ed.)*. Cambridge: Cambridge University Press.
- Husain, N. (2015). *Language and Language Skills*.
- Hymes, D. H. (1972). *On Communicative Competence*. Penguin Education, Penguin Books Ltd.
- IELTS. (2018). Information for Candidates: Introducing IELTS to test takers (pp. 1-8, Publication). [www.ielts.org](https://www.ielts.org/-/media/publications/information-for-candidates/ielts-information-for-candidates-english-uk.ashx). Retrieved from <https://www.ielts.org/-/media/publications/information-for-candidates/ielts-information-for-candidates-english-uk.ashx>.
- Jensen, A. R. (1969). Understanding Readiness: An Occasional Paper. *Clearinghouse on Early Childhood Education*, 1-1.
- Jian-lan, B. (2007). An analysis of authenticity in CET-4 and TEM-8. *Sino-US English Teaching*, 4(2): 28-33. Retrieved 20th January 2008 from <http://www.linguist.org.cn/doc/su200702/su20070206.pdf>.
- Karabulut. A. (2007). *Micro level impacts of foreign language test (university entrance exam) in Turkey: A washback study* ((Yayınlanmamış Yüksek Lisans Tezi). IowaState University. USA.
- Karataş, H., & Fer, S. (2011). CIPP Evaluation Model Scale: Development, Reliability and Validity. *Proc. Soc. Behav. Sci.* 15:592-599.
- Language Training and Testing Center. (2002). *GEPT Advanced Level research report*. Taipei, Taiwan: Author.
- Lewkowicz, J. (1997): *Investigating authenticity in language testing*. Unpublished PhD dissertation, University of Lancaster.
- Leydens, J. & Thompson, D. (1997). Writing rubrics design (EPICS) I, *Internal Communication*, Design (EPICS) Program, Colorado School of Mines.
- Lightbown, P. M., & Spada, N. (2006). *How languages are learned*. Oxford: Oxford University Press.
- Lightbown, P., & Spada, N. M. (2013). *How languages are learned (4th Ed.)*. Oxford: Oxford University Press.

- Linville, H. (2011). *Assessment Principles II and Techniques for Assessment and Testing* Retrieved from <http://blackboard.umbc.edu>
- Longman, P. (n.d.). Longman Dictionary of Contemporary English Online. Retrieved May 5, 2018, from <https://www.ldoceonline.com/>
- Madsen, H. S. (1983). *Techniques in Testing*. Oxford: Oxford University Press,
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276–282.
- McNeill, D. (1966). "Developmental psycholinguistics". Smith, F. & Miller, G. A. (eds.) *The Genesis of Language*. Cambridge, Mass.: M.I.T. Press.
- Messick S. (1989). *Validity*. In Linn, R.L. (Ed.). *Educational measurement (3rd Ed.)*. New York: American Council on Education/Collier Macmillan, pp 13-103.
- Messick, S. (1996). Validity and washback in language testing. *Language testing*, 13(3), 241-256.
- Mitchell, T. R. (1997). Matching motivational strategies with organizational contexts. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 19, pp. 57–149). Greenwich, CT: JAI.
- MONE. (2013). *English Language Curriculum for Primary Education*. Ankara: NEM.
- Moeller, A. K., & Catalano, T. (2015). Foreign language teaching and learning. Faculty Publications: *Department of Teaching, Learning and Teacher Education*, 196, 327-332.
- Moskal, B. M. (2000). "Scoring rubrics: What, when and how?" *Practical Assessment, Research & Evaluation*, 7 (3) [Available Online: <http://pareonline.net/getvn.asp?v=7&n=3>].
- Mousavi, S.A. (2002). *An encyclopaedic dictionary of language testing: (3rd Ed.)*. Taipei: Tung Hua Publications.
- Newby, D. (2011). *Issues in using the EPOSTL*. Newby, D., Fenner, A.-B. & Jones, B. (Eds.) *Using the European portfolio for student teachers of languages*. European Centre for Modern Languages, Council of Europe Publishing.
- O'Grady, W., Dobrovolsky, M., & Katamba, F. (1996). *Contemporary Linguistics: An Introduction*. London: Longman.
- OECD (2005), *Formative Assessment: Improving Learning in Secondary Classrooms*, OECD, Paris.
- Overton, T. (2008). *Assessing learners with special needs: An applied approach (6th Ed.)*. Upper Saddle River, NJ: Prentice Hall

- Oxford University Press. (n.d.). English Dictionary, Thesaurus, & grammar help | Oxford Dictionaries. Retrieved May 5, 2018, from <https://en.oxforddictionaries.com/>
- Oxford, R. (2001) Integrated Skills in the ESL/EFL Classroom. *ERIC Resource Centre*.
- Ozerova, A. (2014). Types of Tests Used in English Language. 1-59. doi: <http://www.bestreferat.ru/referat-39308.html>
- Paterno, A. (1965). "Foreign Language Testing". In Allen, J.P.B. (ed.) *Teaching English as a Second Languages*. New York: McGraw Hill, Inc.
- Patton, M.Q. (1997). *Utilization-Focused Evaluation*. Thousand Oaks, California: Sage Publications.
- Perrenoud, P. (1998). From Formative Evaluation to a Controlled Regulation of Learning Processes. Towards a Wider Conceptual Field. *Assessment in Education*, 5(1) pp.85-102
- Pesetsky D. (1999). *Linguistic universals and Universal Grammar*. Cambridge, MA: MIT Press.
- Polat, M. (2017). *Akademisyenlerin Yabancı Dil Bilgisi Seviye Tespit Sınavı'nın (YDS) İçeriği ve Ket Vurma Etkisine İlişkin Görüşleri* (Yayımlanmamış yüksek lisans tezi). Eğitim Bilimleri Enstitüsü/Eskişehir Osmangazi Üniversitesi.
- Professional Testing Inc. (2006). *Step 9. Conduct the Item Analysis* [PDF]. Professional Testing Inc.
- Qi, L. (2005). Stakeholders' Conflicting Aims Undermine the Washback Function of a High-Stakes Test. *Language Testing*, 22(2), 142–173.
- Rahman, M. M., & Gautam, A. K. (2012). Testing and Evaluation: A Significant Characteristic of Language Learning and Teaching. *LANGUAGE IN INDIA/ Strength for Today and Bright Hope for Tomorrow*, 12(1), 432-442.
- Richards, J. C. (2006). *Communicative Language Teaching Today*. Cambridge: Cambridge University Press.
- Roberts, D.M. (2000). Face validity: Is there a place for this in measurement? *SHIKEN: JALT Test. Eval. SIG Newsl.* 4(2), 5.
- Sadiku, L. M. (2015). The Importance of Four Skills Reading, Speaking, Writing, Listening in a Lesson Hour. *European Journal of Language and Literature*, 1(1), 29-31.

- Sayın, B. A., & Aslan, M. M. (2016). The Negative Effects of Undergraduate Placement Examination of English (LYS-5) on ELT Students in Turkey. *Participatory Educational Research*, 3(1), 30-39.
- Saville, N. (2000). Investigating the Impact of International Language Examinations. *Research Notes*, 2, 2-7.
- Schellekens, P. (2001). *English as a Barrier to Employment, Education & Training*. London: DfES.
- School of Foreign Languages. (2017, October 16). Akdeniz University /School of Foreign Languages. Retrieved June 18, 2018, from <http://ydy.akdeniz.edu.tr/2017-2018-egitim-ogretim-yili-hazirlik-sube-ders-programlari/>
- Sevimli, S. E. (2007). *The Washback Effects of Foreign Language Component of The University Entrance Examination on The Teaching and Learning Context of English Language Groups in Secondary Education (a case study)*. Unpublished master's thesis. Graduate School of Social Sciences/Gaziantep University.
- Shohamy, E., Donitsa, S., & Irit, F. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317.
- Shomoossi, N., & Tavakoli, M. (2010). Authenticity and Authentication in Language Testing: An Operational Perspective. *Modern Journal of Applied Linguistics*, 2(1), 1-16.
- Sim, S., & Rasiyah, R., (2006). Relationship between item difficulty and discrimination indices in true/false- type multiple-choice questions of a para-clinical multidisciplinary paper. *Annals Academy of Medicine*, 35 (2), 67-71
- Sonnentag, S. & Frese, M. (2005). *Performance Concepts and Performance Theory*. 1-25.
- Student Selection and Placement Centre. (2018). OSYM. Retrieved May 05, 2018, from <http://www.osym.gov.tr/TR,8860/hakkinda.html>
- Surgenor, P. (2010). Summative & Formative Assessment. *UCD Teaching and Learning Resources/ Teaching Toolkit*, 1-2.
- Thorndike, E. L. (1977). *The fundamentals of learning*. New York: Columbia University, Teachers College.
- Tuna, A. & Kaçar, A. (2005). İlköğretim matematik öğretmenliği programına başlayan öğrencilerin lise 2 matematik konularındaki hazır bulunuşluk düzeyleri. *Kastamonu Eğitim Dergisi*, 13(1), 117-128.

- Ünal, M. (2005). *Eğitim fakültelerinde ortak ders olarak okutulan yabancı dil derslerinde öğrencilerin bilişsel hazır bulunuşluk düzeylerinin akademik başarıya etkisi*. Yayınlanmamış yüksek lisans tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Valette, R. M. (1977). *Modern Language Testing (2nd Ed.)*. Harcourt Brace Jovanovich.
- Vural, T. (2004). *An Evaluation of the Curriculum Applied at the Preparatory English Classes of Yıldız Technical University*. Unpublished Master's Thesis, Yıldız Technical University, İstanbul.
- W.White, R. (1959). Motivation Reconsidered: The Concept Of Competence. *Psychological review*. 66. 297-333.
- Weinert, F. E. (2001). *Concept of competence: A conceptual clarification*. Ashland, OH, US: Hogrefe & Huber Publishers.
- Weir, C. (2005). *Language testing and validation: an evidence-based approach*. UK: Palgrave Macmillan.
- Weir, C.J. (1990). *Communicative Language Testing*. London: Prentice Hall.
- Widdowson, H. G. (1983). *Learning Purpose and Language Use*. Oxford: Oxford University Press.
- Wither, C. (2017). *Difference between Measurement and Evaluation*." Difference Between.net. November 16, 2017 < <http://www.differencebetween.net/language/difference-between-measurement-and-evaluation/> >.
- Wood, R. (1993). *Assessment and testing*. Cambridge: Cambridge University Press.
- Yenilmez, K. & Kakmacı, Ö. (2008). İlköğretim yedinci sınıf öğrencilerinin matematikteki hazır bulunuşluk düzeyi. *Kastamonu Eğitim Dergisi*, 16(2), 529-542.
- Yıldırım, Ö. (2010). Washback effects of a high-stakes university entrance exam: Effects of the English section of the university entrance exam on future language teachers in Turkey. *The Asian EFL Journal Quarterly*, 12(2), 92-116.
- Yoneda, M. (2012). Designing Assessment Tools: The Principles of Language Assessment. Bulletin of Mukogawa Women's University. *Humanities and Social Science*, 41-49.
- Yüksel, A., Marangoz, D.Ç. & Canaran, N.T. (2004). Farklılaştırılmış öğretim stratejileri. 09.04.2011 tarihinde

<http://www.erg.sabanciuniv.edu/iok2004/bildiriler/Arzu%20Yuksel.doc>
adresinden alınmıştır.



APPENDICES

Appendix A: Test of English Language Proficiency



TEST OF ENGLISH LANGUAGE PROFICIENCY (TELP)

Candidate Name _____

“Dear Participant,

The aim of this test is to determine the score differences of communicative skills between the students who attended the preparatory class and the ones who did not. The scores of this test will be compared with LYS-5 Foreign Language Proficiency Test scores as well. This test consists of four sections including reading, writing, listening, and speaking. The speaking session will be conducted separately at another time. This test contributes to an M.A. thesis at English Language Teaching Department of Akdeniz University. Your answers and the test scores will be kept confidential and will be utilized for research purposes only.”

Thank you for your contribution and time in advance,

Burak ASMA - Research Assistant, Akdeniz University

Table of Content

SECTION	SCORE
Reading Comprehension	
Listening Comprehension	
Writing Competence	
Speaking Competence	
Total	

READING SECTION

Reading Task 1: Passage and Questions

Directions: Read the passage. Then answer the questions. You have 20 minutes to complete this task.

Risk-Taking and the Monkey Economy

- 1) Humans are uniquely smart among all the other species on the planet. We are capable of outstanding feats of technology and engineering. Then why are we so prone to making mistakes? And why do we tend to make the same ones time and time again? When Primate Psychologist Laurie Santos from the Comparative Cognition Lab at Yale University posed this question to her team, they were thinking in particular of the errors of judgement which led to the recent collapse of the financial markets. Santos came to two possible answers to this question. Either humans have designed environments which are too complex for us to fully understand, or we are biologically prone to making bad decisions.
- 2) In order to test these theories, the team selected a group of Brown Capuchin monkeys. Monkeys were selected for the test because, as distant relatives of humans, they are intelligent and have the capacity to learn. However, they are not influenced by any of the technological or cultural environments which affect human decision-making. The team wanted to test whether the capuchin monkeys, when put into similar situations as humans, would make the same mistakes.
- 3) Of particular interest to the scientists was whether monkeys would make the same mistakes when making financial decisions. In order to find out, they had to introduce the monkeys to money. The team distributed metal discs to the monkeys, and taught them that the discs could be exchanged with team-members for food. The monkeys soon cottoned on, and as well as learning simple exchange techniques, were soon able to distinguish 'bargains' – If one team-member offered two grapes in exchange for a metal disc and another team-member offered one grape, the monkeys chose the two-grape option. Interestingly, when the data about the monkey's purchasing strategies was compared with economist's data on human behavior, there was a perfect match.
- 4) So, after establishing that the monkey market was operating effectively, the team decided to introduce some problems which humans generally get wrong. One of these issues is risk-taking. Imagine that someone gave you \$1000. In addition to this \$1000, you can receive either A) an additional \$500 or B) someone tosses a coin and if it lands 'heads' you receive an additional \$1000, but if it lands 'tails' you receive no more money. Of these options, most people tend to choose option A. They prefer guaranteed earnings, rather than running the risk of receiving nothing. Now imagine a second situation in which you are given \$2000. Now, you can choose to either A) lose \$500, leaving you with a total of \$1500, or B) toss a coin; if it lands 'heads' you lose nothing, but if it lands 'tails' you lose \$1000, leaving you with only \$1000. Interestingly, when we stand to lose money, we tend to choose the more risky choice, option B. And as we know from the experience of financial investors and gamblers, it is unwise to take risks when we are on a losing streak.
- 5) So would the monkeys make the same basic error of judgement? The team put them to the test by giving them similar options. In the first test, monkeys had the option of exchanging their disc for one grape and receiving one bonus grape, or exchanging the grape for one grape and sometimes receiving two bonus grapes and sometimes receiving no bonus. Then the experiment was reversed. Monkeys were offered three grapes, but in

option A were only actually given two grapes. In option B, they had a fifty-fifty chance of receiving all three grapes or one grape only. The results were that monkeys, like humans, take more risks in times of loss.

- 6) The implications of this experiment are that because monkeys make the same irrational judgements that humans do, maybe human error is not a result of the complexity of our financial institutions, but is imbedded in our evolutionary history. If this is the case, our errors of judgement will be very difficult to overcome. On a more optimistic note however, humans are fully capable of overcoming limitations once we have identified them. By recognizing them, we can design technologies which will help us to make better choices in future.

Questions:

1. Why did the psychologist "Laurie Santos" carry out the experiment stated in the passage?

- A) To find out whether external factors affect monkeys' decision-making process.
- B) To reveal the similarities between humans and Brown Capuchin monkeys
- C) To question what the source of human mistakes is
- D) To investigate whether there is an exact solution not to make mistakes

2. Which paragraph informs us about the steps followed during the research?

- A) Paragraph 2
- B) Paragraph 3
- C) Paragraph 4
- D) Paragraph 5

3. The word "cottoned on" in the passage is closest in meaning to

- A) achieved
- B) confused
- C) learnt
- D) understood

4. In paragraph 5 of the text, there is a missing sentence. The paragraph is repeated below and shows four letters (A, B, C, and D) that indicate where the following sentence could be added.

"It turned out that monkeys, like humans, chose the less risky option in times of plenty."

Where would the sentence best fit?

So would the monkeys make the same basic error of judgement? The team put them to the test by giving them similar options. (A) In the first test, monkeys had the option of exchanging their disc for one grape and receiving one bonus grape, or exchanging the grape for one grape and sometimes receiving two bonus grapes and sometimes receiving no bonus. (B) Then the experiment was reversed. (C) Monkeys were offered three grapes, but in option A were only actually given two grapes. In option B, they had a fifty-fifty chance of receiving all three grapes or one grape only. (D) The results were that monkeys, like humans, take more risks in times of loss.

5. In paragraph 6, "them" refers to

- A) implications
- B) financial situations
- C) choices
- D) limitations

6. All of the following questions about Brown Capuchin monkeys and humans are answered in this text EXCEPT:

- A) Why are humans so prone to making mistakes?
- B) Why did the monkeys prefer the more risky options?
- C) Why did the researchers conduct this experiment?
- D) Why were the Brown Capuchin monkeys involved in the experiment?

Reading Task 2: Passage and Questions

Directions: Read the passage. Then answer the questions. You have 20 minutes to complete this task.

Social housing in Britain

- 1) During the past 20 years in Britain there has been a significant decrease in the number of social homes in the housing stock, down from 5.3m to 4.8m. The proportion of social housing has fallen from 29% to 18% during the same period. This is largely due to the policies of Margaret Thatcher's government during the 1980's which forced local councils to sell homes under market price to existing tenants under a 'right to buy' scheme and prevented them from building new houses. New social homes were then to be paid for by central government and managed by local housing associations.
 - 2) Next month, the government is expected to announce a significant increase in the Social Housing Department's £1.7 billion annual budget and also intends to make the application process for social housing simpler. The additional £2 billion will build about 50,000 new houses each year at current building costs. Still more houses could be built if subsidies were reduced.
 - 3) The UK government is hoping that the extra investment will improve the housing situation. Britain with her increasing population has built fewer new houses than are needed, with a shortfall of 100,000 a year according to Shelter, a housing charity. The result is a boom in house prices that has made owning a home unaffordable for many, especially in London and the south of England. Key public sector workers, such as nurses and teachers, are among those affected.
 - 4) In order to increase the social housing stock the government is using a process known as planning gain. Town councils are increasing the amount of social housing developers must build as part of a new building project and which they must give to the local housing association. Even without the financial support of central the government, some local councils in England are using planning gain to increase the proportion of social housing stock. In expensive Cambridge, the council wants 25% of new housing to be social; the figure is 35% in Bristol, while Manchester is planning 40% over the next twenty years.
 - 5) Will this housing policy create new sink estates? Hopefully, not. Housing planners have learnt from the mistakes of the 1960s and 1970s when large council housing estates were constructed. Builders have got better at design and planning mixed-use developments where social housing is mixed with, and indistinguishable from, private housing. Social housing developments are winning design awards - a project in London won the Housing Design Award — though it is true that some council estates that now illustrate some of the worst aspects of 1960s architecture won awards at the time.
 - 6) The management of social housing stock has largely moved from local councils to housing associations. Housing associations look after the maintenance of the existing housing stock, getting repairs done and dealing with problems like prostitution and drugs while employing estate security and on-site maintenance staff. One significant change is that planners have learned to build smaller housing developments.
-

7) The significant drawback of social housing still remains: it discourages mobility. What happens to the nurse who lives in cheap social housing in one town, and is offered a job in a region that does not provide her with new social housing? The government wants to encourage initiative but is providing a housing system that makes it difficult for people to change their lives. Public-sector workers are increasingly being priced out of London and other expensive parts of the country and, as a result, are unable to take advantage of opportunities available to them.

A. Match each heading to the most suitable paragraph, there is an extra heading.

- a. Supposed growth
- b. Developers are asked for renewals
- c. Learning through the past
- d. The cost of changing home
- e. Ongoing handicaps of social housing
- f. A decline in the quantity
- g. Officials of state sector are excluded
- h. From local housing to associations

- | | |
|----------------------|----------------------|
| 1. Paragraph 1 | 5. Paragraph 5 |
| 2. Paragraph 2 | 6. Paragraph 6 |
| 3. Paragraph 3 | 7. Paragraph 7 |
| 4. Paragraph 4 | |

B. Select *True, False, or Not given*

- 8. The government of 1980s promoted the construction of new houses.
- 9. An easier application procedure is supposed to be suggested by the government.
- 10. There is a growing effort which is being spent on promoting locals to buy new houses via social housing.
- 11. Current social housing councils do their best to be awarded.
- 12. Recently, administration of social housing has made progress in terms of providing support for the finalized houses.
- 13. The recent progress on social housing overcame the problems.

LISTENING SECTION

Name _____

Instructions to candidates

Do not open this question paper until you are told to do so.

Write your name in the spaces at the top of this page.

Listen to the instructions for each task carefully.

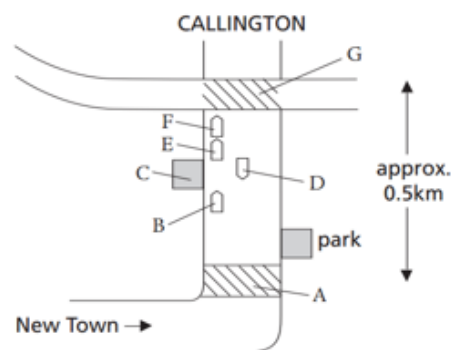
Answer all the questions.

Information for candidates

There are **two** tasks in the test. You will hear each part once only. There are **15** questions. Each question carries one point. For each task of the test, there will be time for you to look through the questions and time for you to check your answers.

Task 1: Complete the form below, using **NO MORE THAN THREE WORDS AND/OR NUMBER** for each answer.

TRAFFIC INSURANCE POLICY FORM	
Department in charge:	(1)..... Insurance
Customer details:	
Name:	Elisabeth (2)
Date of birth:	8.10.1975
Address:	(3)..... (street) Callington (town)
Policy number: (4)
Accident details:	
Date: (5)
Time:	Approx..... (6)
Supporting evidence:(7)



(8). railway bridge

(9). blue van

Task 2: Choose the correct letter, A,B,C or D.

11. What do male elephants do when they wish to become a pair?
- A) join new elephant groups
 - B) make subvocal sounds
 - C) go for a long walk
 - D) gather in small groups
12. What behaviour of the elephants aroused scientists' attention?
- A) communication over long distances
 - B) co-ordination of separate families
 - C) utterance of odd sounds during greeting
 - D) keen sense of danger from outside
13. Which sense enables elephants to communicate across long distances?
- A) hearing
 - B) smelling
 - C) touching
 - D) tasting
14. What hindered the team of scientists to hear the sounds the elephants produced?
- A) the place where the experiment conducted
 - B) infrasonic structure of the sounds
 - C) an unexpected loudspeaker fault
 - D) pauses in broadcast network
15. How many types of call occur before mating activity?
- A) one
 - B) two
 - C) three
 - D) four
-

WRITING SECTION

Task 1

Read the passage below and summarize it using one sentence. You have 10 minutes to finish this task. Your response will be judged on the quality of your writing and on how well your response presents the key points in the passage. Make sure to include the main points of the reading passage in a full, single sentence of no more than 25 words.

The development of allergy is a result of a complex interaction between genetic and many environmental factors that may protect against or promote its development. Factors such as pollution have also been linked to the increased prevalence of allergic diseases during childhood in developed countries.

“Exposure to many environmental factors have changed during the last decades, including exposure to tobacco smoke, which appears to increase the risk for airway infections and asthma. Besides, exposure to allergy developing agents such as food, house dust mites, pets and pollens is a prerequisite for development of allergic diseases, but also many other unknown factors may play a role,” explains Prof. Halken.

“Genetic factors may also influence the susceptibility to different environmental factors and also influences the pattern of symptoms of the individual child,” says Prof. Halken. “Some studies report that a child with atopic dermatitis and a family background show higher risk to develop asthma later in life. When one of the parents is allergic, the child is predisposed to be allergic, and the is even higher when both parents suffer this disease.”

The expression of allergic disease may vary with age, and some symptoms may disappear being replaced by other atopic dermatitis, gastrointestinal symptoms and recurrent wheezing, whereas bronchial asthma and allergic rhinoconjunctivitis are the main allergic symptoms in childhood.” In that sense, allergic reactions to foods, mainly cow’s milk protein, are the commonest manifestation in the first years of life, whereas allergy to inhalant agents mostly occurs later in childhood.

Physicians agree that one of the key elements in achieving better management of childhood allergy is to improve diagnostic techniques and to develop treatments that do not only reduce symptoms, but can induce a permanent cure. “Early diagnosis can lead to effective treatment to reduce symptoms and improve quality of life. Knowledge about allergies can help patients to avoid contact with offending agents, and thereby to reduce symptoms and avoid risky situations, which may even be life threatening,” points out Prof. Halken.

Answer:

.....
.....

SPEAKING SECTION

Question 1: Read a text aloud

Directions: In this part of the test, you will read aloud the text below. You will have 45 seconds to prepare. Then you will have 45 seconds to read the text aloud.

Forests are the lifeguards of the snowy peaks of the Alps. They provide a natural barrier against avalanches and landslides, but the skiing industry, which proved a boon for poor Alpine farmers, is damaging the environment. Forests have been felled to make way for more ski runs, car parks, and hotels, and Alpine meadows have been abandoned by farmers keen to exploit tourism. Consequently, the avalanche has now become a common phenomenon. Forestry experts estimate that two-thirds of the several thousand avalanches that descend into inhabited parts each year are the result of forest depletion.

Question 2: Describe a picture

Directions: In this part of the test, you will describe the picture below in as much detail as you can. You will have 30 seconds to prepare your response. Then you will have 45 seconds to speak about the picture.



Questions 3–5: Respond to questions

Directions: In this part of the test, you will answer three questions. For each question, begin responding immediately after you hear a beep. No preparation time is provided. You will have 15 seconds to respond to questions 3 and 4, and 30 seconds to respond to Question 5.

(Narrator): Imagine that an Australian company is carrying out a research in your country. You volunteered to take part in a telephone interview about reading book.

Question 3: How often do you read a book?

Question 4: What kinds of books do you usually read?

Question 5: Describe your favorite book.

Question 6: Express an opinion

Directions: In this part of the test, you will give your opinion about a specific topic. Be sure to say as much as you can in the time allowed. You will have 15 seconds to prepare. Then you will have 60 seconds to speak.

Question:

Some people prefer to study at home rather than to go to the school. They believe it is more time-efficient and easier. What is your opinion about studying at home instead of going to the school? Give reasons for your opinion.

Appendix B / Scoring Rubric for pronunciation

Score	RESPONSE DESCRIPTION
3	Pronunciation is highly intelligible, though the response may include minor lapses and/or other language influence.
2	Pronunciation is generally intelligible, though it includes some lapses and/or other language influence.
1	Pronunciation may be intelligible at times, but significant other language influence interferes with appropriate delivery of the text.
0	No response OR no English in the response OR response is completely unrelated to the test.

Appendix C/ Scoring Rubric for intonation and stress

Score	RESPONSE DESCRIPTION
3	Use of emphases, pauses, and rising and falling pitch is appropriate to the text.
2	Use of emphases, pauses, and rising and falling pitch is generally appropriate to the text, though the response includes some lapses and/or moderate other language influence.
1	Use of emphases, pauses, and rising and falling pitch is not appropriate, and the response includes significant other language influence.
0	No response OR no English in the response OR the response is completely unrelated to the test.

Appendix D/ Scoring Rubric for describing a picture

Score	RESPONSE DESCRIPTION
3	The response describes the main features of the picture. -The delivery may require some listener effort, but it is generally intelligible. -The choice of vocabulary and use of structures allows coherent expression of ideas.
2	The response is connected to the picture, but meaning may be obscured in places. -The delivery requires some listener effort. -The choice of vocabulary and use of structures may be limited and may interfere with overall comprehensibility.
1	The response may be connected to the picture, but the speaker's ability to produce intelligible language is severely limited.

	<p>-The delivery may require significant listener effort.</p> <p>-The choice of vocabulary and use of structures is severely limited OR significantly interferes with comprehensibility.</p>
0	No response OR no English in the response OR the response is completely unrelated to the test.

Appendix E/ Scoring Rubric for responding questions (3)

Score	RESPONSE DESCRIPTION
3	<p>The response is a full, relevant, socially appropriate reply to the question. In the case of Questions 4-5, information from the prompt is accurate.</p> <p>-The delivery requires little listener effort.</p> <p>-The choice of vocabulary is appropriate.</p> <p>The use of structures fulfils the demands of the task.</p>
2	<p>The response is a partially effective reply to the question, but is not complete, fully appropriate, or in the case of Questions 4-5, fully accurate.</p> <p>-The delivery may require some listener effort but is mostly intelligible.</p> <p>-The choice of vocabulary may be limited or somewhat inexact, although overall meaning is clear.</p> <p>-The use of structures may require some listener effort for interpretation.</p> <p>-In the case of Question 4-5, the speaker may locate the relevant information in the prompt but fail to distinguish it from irrelevant information or fail to transform the written language so a listener can easily understand it.</p>
1	<p>The response does not answer the question effectively. Relevant information is not conveyed successfully.</p> <p>-The delivery may impede or prevent listener comprehension.</p> <p>-The choice of vocabulary may be inaccurate or rely on repetition of the prompt.</p> <p>-The use of structures may interfere with comprehensibility.</p>
0	No response OR no English in the response OR the response is completely unrelated to the test.

Appendix F/ Scoring Rubric for expressing an opinion

Score	RESPONSE DESCRIPTION
5	<p>The response clearly indicates the speaker's choice or opinion, and support of the choice or opinion is readily intelligible, sustained, and coherent.</p> <p>The response is characterized by ALL of the following:</p> <ul style="list-style-type: none"> -The speaker's choice or opinion is supported with reason(s), details, arguments, or exemplifications; relationships between ideas are clear. -The speech is clear with generally well-paced flow. It may include minor lapses or minor difficulties with pronunciation or intonation patterns that do not affect overall intelligibility. -Good control of basic and complex structures, as appropriate, is exhibited. Some minor errors may be noticeable but they do not obscure meaning. -The use of vocabulary is effective, with allowance for occasional minor accuracy.
4	<p>The response clearly indicates the speaker's choice or opinion and adequately supports or develops the choice or opinion.</p> <ul style="list-style-type: none"> -The response explains the reason(s) for the speaker's choice or opinion, although the explanation may not be fully developed; relationships between ideas are mostly clear, with occasional lapses. -Minor difficulties with pronunciation, intonation, or pacing are noticeable and may require listener effort at times, although overall intelligibility is not significantly affected. -The response demonstrates fairly automatic and effective use of grammar but may be somewhat limited in the range of structures used. -The use of vocabulary is fairly effective. Some vocabulary may be inaccurate or imprecise.
3	<p>The response expresses a choice, preference, or opinion, but development and support of the choice or opinion is limited.</p> <ul style="list-style-type: none"> -The response provides at least one reason supporting the choice, preference, or opinion. However, it provides little or no elaboration of the reason, repeats itself with no new information, is vague, or is unclear. -The speech is basically intelligible though listener effort may be needed because of unclear articulation, awkward intonation, or choppy of grammar; for the most part, only basic sentence structures are used successfully. -The use of vocabulary is limited.
2	<p>The response states a choice, preference, or opinion relevant to the prompt, but support for the choice, preference, or opinion is missing, unintelligible, or incoherent.</p> <ul style="list-style-type: none"> -Consistent difficulties with pronunciation, stress, and intonation cause considerable listener effort; delivery is choppy, fragmented, or telegraphic; there may be long pauses and frequent hesitations. -Control of grammar severely limits expression of ideas and clarity of connections among ideas. -The use of vocabulary is severely limited or highly repetitious.
1	<p>The response is limited to reading the prompt or the directions aloud OR the response fails to state and intelligible choice, preference, or opinion as required by the prompt OR the response consists of isolated words or phrases, or mixtures of the first language and English.</p>
0	<p>No response OR no English in the response OR the response is completely unrelated to the test.</p>

Appendix G/ Scoring Rubric for summarizing a Text

Enabling skills and other traits scored	Content: 2 Provides a good summary of the text. All relevant aspects mentioned 1 Provides a fair summary of the text but misses one or two aspects 0 Omits or misrepresents the main aspects of the text
	Form: 1 Is written in one, single, complete sentence 0 Not written in one, single, complete sentence or contains fewer than 5 or more than 75 words. Summary is written in capital letters
	Grammar: 2 Has correct grammatical structure 1 Contains grammatical errors but with no hindrance to communication 0 Has defective grammatical structure which could hinder communication
	Vocabulary: 2 Has appropriate choice of words 1 Contains lexical errors but with no hindrance to communication 0 Has defective word choice which could hinder communication

Appendix J/ Scoring Rubric for writing an Essay

<p>Enabling skills and other traits scored</p>	<p>Content: 3 Adequately deals with the prompt 2 Deals with the prompt but does not deal with one minor aspect 1 Deals with the prompt but omits a major aspect or more than one minor aspect 0 Does not deal properly with the prompt</p>
	<p>Form: 2 Length is between 200 and 300 words 1 Length is between 120 and 199 or between 301 and 380 words 0 Length is less than 120 or more than 380 words. Essay is written in capital letters, contains no punctuation or only consists of bullet points or very short sentences</p>
	<p>Development, structure and coherence: 2 Shows good development and logical structure 1 Is incidentally less well structured, and some elements or paragraphs are poorly linked 0 Lacks coherence and mainly consists of lists or loose elements</p>
	<p>Grammar: 2 Shows consistent grammatical control of complex language. Errors are rare and difficult to spot 1 Shows a relatively high degree of grammatical control. No mistakes which would lead to misunderstandings 0 Contains mainly simple structures and/or several basic mistakes</p>
	<p>General linguistic range: 2 Exhibits smooth mastery of a wide range of language to formulate thoughts precisely, give emphasis, differentiate and eliminate ambiguity. No sign that the test taker is restricted in what they want to communicate 1 Sufficient range of language to provide clear descriptions, express viewpoints and develop arguments 0 Contains mainly basic language and lacks precision</p>
	<p>Vocabulary range: 2 Good command of a broad lexical repertoire, idiomatic expressions and colloquialisms 1 Shows a good range of vocabulary for matters connected to general academic topics. Lexical shortcomings lead to circumlocution or some imprecision 0 Contains mainly basic vocabulary insufficient to deal with the topic at the required level</p>
	<p>Spelling: 2 Correct spelling 1 One spelling error 0 More than one spelling error</p>

Appendix K/ LYS-5 English Test scores of the participants

PARTICIPANTS	SCORES	PARTICIPANTS	SCORES
P1	97,5	P42	83,03
P2	96,25	P43	83,03
P3	93,75	P44	82,5
P4	92,5	P45	82,5
P5	91,25	P46	82,5
P6	91,25	P47	82,5
P7	90	P48	82,5
P8	90	P49	81,25
P9	90	P50	81,25
P10	88,75	P51	81,25
P11	88,75	P52	81,25
P12	88,75	P53	81,25
P13	88,75	P54	81,25
P14	88,75	P55	80
P15	88,75	P56	80
P16	88,75	P57	80
P17	87,5	P58	78,75
P18	87,5	P59	78,75
P19	87,5	P60	78,75
P20	87,5	P61	78,75
P21	87,5	P62	78,75
P22	86,25	P63	78,75
P23	86,25	P64	77,5
P24	86,25	P65	77,5
P25	86,25	P66	77,5
P26	86,25	P67	77,5
P27	85	P68	77,5
P28	85	P69	77,5
P29	85	P70	77,5
P30	85	P71	77,5
P31	85	P72	77,5
P32	85	P73	77,5
P33	85	P74	77,5
P34	85	P75	77,5
P35	83,75	P76	76,25
P36	83,75	P77	75
P37	83,75	P78	73,75
P38	83,75	P79	72,5
P39	83,75	P80	68,75
P40	83,75	P81	67,5
P41	83,03		

CURRICULUM VITAE

Burak ASMA

Foreign Language Education

Education

Bachelor of Arts : Akdeniz University, Faculty of Education, Department of Foreign Language Education, Division of English Language Teaching

Master of Arts : Akdeniz University, Graduate School of Educational Sciences, Department of Foreign Language Education, Program of English Language Teaching (Master's with thesis)

Work Experience

2015-2018 Research Assistant, Akdeniz University, Graduate School of Educational Sciences

Publications (Published in Proceedings)

Asma B., Ünal O., "Türk Üniversitelerinin Uluslararası Derecelendirme Kuruluşlarındaki Mevcut Durumu ", V. Uluslararası Avrasya Eğitim Araştırmaları Kongresi, Antalya, Türkiye, 2-5 Mayıs 2018, pp.32-32

Direkci B., Asma B., Şimşek B., "Developing a Scale for Students' Attitudes towards their Mother-tongue", 3. International Symposium on Language Education and Teaching, Roma, İtalya, 20-23 Nisan 2017, pp.80-81

Direkci B., Şimşek B., Asma B., "Examining Internationalization Status of Scientific Papers in Turkey: Akdeniz University Case", 8th International

Graduate Education Symposium, Lefkoşa, Kuzey Kıbrıs Türk Cum., 11-12 Mayıs 2017, pp.1-1

Caner M. , Asma B., "Investigating Pre-Service Teachers' Reflections on Teacher Development Courses", Uluslararası Multidisipliner Çalışmaları Kongresi, Antalya, Türkiye, 25-26 Kasım 2017, pp.5-5

Asma B., Saraç H.S., "A Study of Historical Views of Political Parties on Foreign Language Teaching", CUELT Çukurova International ELT Teachers Conferences "Replacing Cliches: Teaching Subordinated to Learning", Adana, Türkiye, 28-29 Nisan 2016, pp.40-40

Asma B., "An Analysis of the role of optimal arousal, novelty and learner competency in EFL Learners' motivation levels", CUELT Çukurova International ELT Teachers Conferences "Replacing Cliches: Teaching Subordinated to Learning", Adana, Türkiye, 28-29 Nisan 2016, pp.30-30

Personal Information

Date and Place of Birth : 10.07.1992 / Kavak (SAMSUN)
Spoken Foreign Languages : English, German

Contact Information

e-mail : asmaburak@gmail.com

BİLDİRİM

Hazırladığım tezin tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin kâğıt ve elektronik kopyalarının Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

Tezimin 1 (bir) yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin tamamı her yerden erişime açılabilir.

27.07.2018

Burak ASMA



Tez Orijinallik Raporu

AKDENİZ UNIVERSITY
INSTITUTE OF EDUCATIONAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGE EDUCATION
PROGRAM OF ENGLISH LANGUAGE TEACHING

EXAMINING THE RELATIONSHIP BETWEEN ENGLISH ADMISSION
TEST (EYS-5) AND FOUR LANGUAGE SKILLS OF
UNDERGRADUATE STUDENTS

MASTER'S THESIS
Doç. Dr. ASMA

Eğleşmeler Genel Bakış		
%12		
1	Yazma ve Okuma	%1
2	Yazma ve Okuma	%1
3	Yazma ve Okuma	%1
4	Yazma ve Okuma	%1
5	Yazma ve Okuma	%1
6	Yazma ve Okuma	%1
7	Yazma ve Okuma	%1
8	Yazma ve Okuma	%1
9	Yazma ve Okuma	%1
10	Yazma ve Okuma	%1
11	Yazma ve Okuma	%1
12	Yazma ve Okuma	%1
13	Yazma ve Okuma	%1

Dr. Öğr. Ü.F. Özlem Saka