

FOR REFERENCE

NOT TO BE TAKEN FROM THIS ROOM

Evaluation of the Boğaziçi University English Proficiency Test

Thesis submitted to the
Institute of Social Sciences
in partial satisfaction of the requirements for the degree of

Master of Arts
in
Foreign Language Education

by
Çiler Hatipoğlu

Bogazici University Library



39001100372302

14

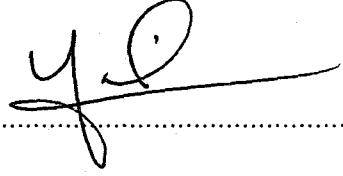
Boğaziçi University

1999

The thesis of Çiler Hatipoğlu

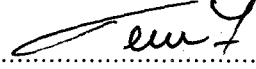
is approved by

Assist. Prof. Dr. Yasemin Bayyurt

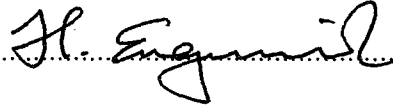


(Committee Chairperson)

Prof. Dr. Cem Alptekin



Prof. Dr. Hüsnü Enginarlar



BOĞAZİÇİ
ÜNİVERSİTESİ
KÜTÜPHANESİ



349785

August 1999



ACKNOWLEDGEMENTS

First and foremost I would like to express my sincere gratitude to my thesis adviser Assist. Prof. Dr. Yasemin Bayyurt, for her patience and efforts to push me to the realms of independence, and her respect in my work treating me like a colleague rather than an MA students while doing this thesis.

I would like to thank Prof. Dr. Cem Alptekin, for the time he has devoted in providing advice and guidance at all stages throughout the preparation of this thesis.

I would also like to thank Mrs. Eileen Yalçın for the time she has spent on editing my thesis and most importantly training me as an English teacher. I also appreciate the time Dr. Judy Doyum has devoted to reading and commenting on my thesis.

I wish to thank Prof. Dr. Rıfat Okçabol for his help in my statistical analyses in getting me in the right track.

I am indebted to my dear friends Elif Öz, Evrim Eveyik and Hande Kür for the endless support, and understanding they have given to me. Through the bad times they have been the best reasons for me to carry on.

I would also like to thank my family for their encouragement not only with this thesis but also with much else besides.

Finally, my deepest gratitude goes to my loving husband İlhan, who has provided me with all the encouragement I needed and has always been there to share my emotions, to live with my troubles and my success as if they were his own.

ABSTRACT

Evaluation of the Boğaziçi University English Proficiency Test

by
Çiler Hatipoğlu

The present study concerns itself with the predictive and face validities of the Boğaziçi University English Proficiency Test (BUEPT), designed to screen students entering Boğaziçi University (BU) for deficiencies in English language skills, which might impede their progress in undergraduate studies. More specifically, the study investigates the following research questions:

1. Is there a relationship between the time students spend in YADYOK and their success on BUEPT?
2. Is BUEPT a valid predictor of the first semester, second semester and overall freshman year academic success of BU students in Foreign Language Education (FLED), Western Languages and Literatures (LL) and Translation and Interpretation (TRANS) Departments? Why or why not?
3. How do the students in FLED, LL and TRANS Departments, evaluate BUEPT as a valid reflector of their level of academic English proficiency and as a predictor of their future academic success?

A total number of 422 (332 female and 89 male) BU students from FLED (207), LL (114) and TRANS (101) Departments, who were admitted to the university between 1992 and 1997, participated in this study.

A quantitative set of data comprised of the BUEPT scores and the first semester, second semester and overall freshman year GPAs of all the subjects in the study was obtained from the BU Registrar's Office. A questionnaire administered to all of the 422 subjects supplied the qualitative data of the study. Descriptive statistics, correlational and key-word analyses were used to analyze the available data.

The results indicated that there was a positive relationship between the time students spent in YADYOK and their exam grades. That is, the longer the students attended prep classes, the more successful they were on BUEPT. However, the BUEPT score was not found to be an effective predictor of the first semester, second semester and overall freshman year academic success, as measured by GPAs, for this group of students. Moreover, the questionnaire data revealed that the majority of the students (78%) perceived the exam neither as a valid reflector of their level of academic English proficiency nor as a good predictor of their future academic performance.

KISA ÖZET

Bu araştırmanın amacı, İngilizce seviyelerinde – üniversite eğitimlerinin ilk yılındaki akademik başarılarını engelleyebilecek derecede - eksiklikleri olan öğrencileri elemek için düzenlenen Boğaziçi Üniversitesi İngilizce Yeterlilik Sınavının (BUEPT) öndeyici ve yüzeysel geçerliliklerini değerlendirmektir. Çalışma şu sorulara cevap aramaktadır: 1) Öğrencilerin YADYOK'taki öğrenim süreleri ile BUEPT'teki başarıları arasında bir ilişki var mıdır? 2) BUEPT, Boğaziçi Üniversitesinin Yabancı Diller Eğitimi (FLED), Batı Dilleri ve Edebiyatı (LL) ve Mütercim Tercümanlık (TRANS) Bölümlerinde öğrenim gören öğrencilerin birinci sömestr, ikinci sömestr ve birinci yıllarındaki akademik başarılarını belirleyen geçerli bir öndeyici midir? 3) Boğaziçi Üniversitesinin, FLED, LL ve TRANS bölümlerinde okuyan öğrenciler, akademik İngilizce seviyelerinin yeterliliğini ölçmede geçerli bir yansıtıcı ve gelecekteki akademik başarılarının bir öndeyicisi olarak BUEPT'i nasıl değerlendiriyorlar?

Bu çalışmaya 1992-1997 yılları arasında Boğaziçi Üniversitesine kabul edilen toplam 422 öğrenci katılmıştır. Örneklem 332'si kız, 89'u erkek öğrencidir. Yine bu öğrencilerin 207'si FLED, 114'ü LL ve 111'i de TRANS bölümlerinde öğrenimlerini sürdürmektedirler.

Çalışmada kullanılan ve öğrencilerin BUEPT skorları, birinci ve ikinci sömestr ile birinci sınıf genel not ortalamalarından oluşan nicel veri grubu, Boğaziçi Üniversitesi Kayıt İşleri'nden temin edilmiştir. Örneklemdeki 422 kişiye doldurtulan bir anket

alışmanın kalitatif veri grubunu meydana getirmiştir. Betimleyici istatistik ile korelasyonel ve anahtar-kelime metotları mevcut verilerin analizi için kullanılmıştır.

Elde edilen bulgular, öğrencilerin YADYOK'taki öğrenim süreleri ile BUEPT skorları arasındaki bağlantının doğru orantılı olduğunu göstermiştir. Diğer bir deyişle, hazırlık sınıflarında daha uzun süre öğrenim gören öğrenciler, BUEPT'te daha başarılı olmuşlardır. Ancak, genel not ortalamaları incelendiğinde, BUEPT'in FLED, LL ve TRANS öğrencilerinin birinci s0mestr, ikinci s0mestr ve birinci yıllarındaki akademik başarılarını belirleyen geçerli bir 3nde yici olmadığı gözlemlenmiştir. Ayrıca, bu bölümlerde öğrenim gören öğrencilerin çoğu (78%), görüşlerini BUEPT'in akademik İngilizce seviyelerinin yeterliliğini ölçen geçerli bir yansıtıcı olmadığı ve gelecekteki akademik başarılarının bir 3n de yicisi olmadığı yönünde belirtmişlerdir.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iii
ABSTRACT	v
KISA ÖZET	vii
LIST OF TABLES AND FIGURES	xii
LIST OF ABBREVIATIONS CONTAINED IN THE TEXT	xiv
CHAPTER 1 INTRODUCTION AND LITERATURE REVIEW	
1.1 Introduction	1
1.1.1 The Psychometric-Structuralist Approach	3
1.1.2 The Psycholinguistic-Sociolinguistic Approach	7
1.1.3 The Communicative Approach	11
1.1.4 The General Proficiency Approach	18
1.2 Validity	21
1.2.1 Predictive Validity	22
1.2.1.1 Positive Conclusions	23
1.2.1.2 Negative Conclusions	24
1.2.1.3 Mixed Conclusions	32
1.2.2 Face Validity	35
CHAPTER 2 BOĞAZIÇI UNIVERSITY ENGLISH PROFICIENCY	
TEST (BUEPT)	39
2.1 Boğaziçi University English Proficiency Test (BUEPT)	39
2.1.1 Background	40
2.1.2 The New Boğaziçi University English Proficiency Test (BUEPT)	44

2.1.3 Comments on the Components of BUEPT	49
2.2 Review of BUEPT and YADYOK Studies	52
CHAPTER 3 METHODOLOGY	55
3.1 Purpose of the Study	55
3.2 Subjects	57
3.3 Data Collection	59
3.4 Piloting	61
3.5 Data Analysis	62
CHAPTER 4 RESULTS AND INTERPRETATIONS	64
PART A: The Answer to Research Questions 1 and 2	64
4.1 Distribution of BUEPT Scores According to the Time Students Spent in the English Preparatory Division (YADYOK) of Boğaziçi University	65
4.2 Statistical Analysis of the Relationship between the BUEPT Scores and the First Semester, Second Semester and Overall Freshman Year GPAs	67
4.2.1 Descriptive Statistics/ Comparison of Means	67
4.2.2 Spearman Rank Order Correlation Coefficients	71
4.3 Within Group Analysis	72
4.3.1 Contingency Tables	72
4.3.1.1 First Semester Contingency Table	73
4.3.1.2 Second Semester Contingency Table	75
4.3.1.3 Freshman Year Contingency Table	77
4.3.2 Comparison of Means within Groups	79
4.3.2.1 Multiple Comparison of the First Semester Means	80

4.3.2.2 Multiple Comparison of the Second Semester Means	81
4.3.2.3 Multiple Comparison of the Freshman Year Means	82
PART B: The Answer to Research Questions 3	84
4.4 Students' Questionnaire Answers	84
4.4.1 Analysis of the First Question	84
4.4.2 Analysis of the Second Question	85
4.4.3 Analysis of the Third Question	87
4.4.4 Analysis of the Fourth Question	90
4.4.5 Analysis of the Fifth Question	98
CHAPTER 5 DISCUSSION AND CONCLUSIONS	102
REFERENCES	115
APPENDIX A	132
APPENDIX B	136
APPENDIX C	140
APPENDIX D	144
APPENDIX E	148
APPENDIX F	160
APPENDIX G	162
APPENDIX H	165

LIST OF TABLES AND FIGURES

Table 3.1.	Reconciled BUEPT Scores	63
Figure 4.1.	Distribution of BUEPT scores according to the time students spent in YADYOK	66
Table 4.1.	Distribution of BUEPT scores according to the time students spent in YADYOK	66
Table 4.2.	Descriptive Statistics	68
Table 4.3.	ANOVA – Scheffe - Comparison of Means	69
Table 4.4.	Spearman Rank Order Correlation Coefficients	71
Table 4.5.	Contingency Table – FIRST SEMESTER	73
Table 4.6.	Contingency Table – SECOND SEMESTER	75
Table 4.7.	Contingency Table –END OF THE FRESHMAN YEAR	77
Table 4.8.	Multiple comparison – ANOVA – Scheffe – FIRST SEMESTER	80
Table 4.9.	Multiple comparison – ANOVA – Scheffe – SECOND SEMESTER	81
Table 4.10.	Multiple comparison – ANOVA – Scheffe – END OF THE FRESHMAN YEAR	82
Table 4.11.	Parts of BUEPT according to their difficulty level	84
Table 4.12.	What is the main goal of prep classes?	85
Figure 4.2.	Should there be an exam measuring your level of academic English proficiency before you begin your freshman year studies? Distribution of students' answers according to the time they spent in YADYOK	87

Table 4.13.	Should there be an exam measuring your level of academic English proficiency before you begin your freshman year studies? WHY DO YOU THINK SO?	88
Figure 4.3.	Did BUEPT accurately reflect the level of your academic English proficiency? Distribution of students' answers according to the time they spent in YADYOK	90
Figure 4.4.	Did BUEPT accurately reflect the level of your academic English proficiency? Distribution of students' answers according to their BUEPT passing grade	92
Table 4.14.	Did BUEPT accurately reflect the level of your academic English proficiency? WHY DO YOU THINK SO?	93
Figure 4.5.	Did BUEPT properly predict your freshman year academic success? Distribution of students' answers according to the time they spent in YADYOK	98
Figure 4.6.	Did BUEPT properly predict your freshman year academic success? Distribution of students' answers according to their BUEPT passing grade	99
Table 4.15.	Did BUEPT properly predict your freshman year academic success? WHY DO YOU THINK SO?	100

LIST OF ABBREVIATIONS CONTAINED IN THE TEXT

ANOVA	= Analysis of Variance
APA	= American Psychological Association
BU	= Boğaziçi University
BUEPT	= Boğaziçi University English Proficiency Test
BÜMED	= Boğaziçi University Alumni Association
CELT	= Cambridge University English Language Test
DELBA	= Davies English Language Battery
EAP	= English for Academic Purposes
EFL	= English as a Foreign Language
ELAT	= English Language Aptitude Test
ELBA	= English Language Battery
ELTS	= English Language Testing System Test
EPT	= English Proficiency Test
EPTB	= English Proficiency Test Battery
ESL	= English as a Second Language
FLED	= Foreign Language Education Department
G1	= Group that passed the BUEPT at first try
G2	= Group that passed the BUEPT after a semester of prep instruction
G3	= Group that passed the BUEPT after two semesters of prep instruction
G	= Group that includes all subjects in the study
GPA	= Grade Point Average
GRE-Q	= Graduate Record Exam of Quantitative Ability
GRE-V	= Graduate Record Exam of Verbal Ability

IELTS	= The International English Language Testing System
I-TOEFL	= Institutional Test of English as a Foreign Language
KELT	= Key English Language Test
LL	= Western Languages and Literatures Department
METU	= Middle East Technical University
MTELP	= Michigan Test of English Language Proficiency
OOEFL	= Oxford Examination in English as a Foreign Language
ÖYS	= University Placement Test (Öğrenci Yerleştirme Sınavı)
ÖSYM	= Student Selection and Placement Center (Öğrenci Seçme ve Yerleştirme Merkezi)
SAT-V	= The Verbal Ability Section of the Scholastic Aptitude Test
SPA	= Semester Point Average
SPSS	= The Statistical Package for Social Sciences
SUNYA	= The State University of New York at Albany
TOEFL	= Test of English as a Foreign Language
TRANS	= Department of Translation and Interpretation
UCLA	= University of California, Los Angeles
YADYOK =	English Preparatory Division of Boğaziçi University (Yabancı Diller Yüksek Okulu)

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Causes of the academic failure of nonnative English speakers in English-medium universities have long concerned researchers in education. This is due to the fact that detailed knowledge of the important determinants of performance can aid in developing remedial tools that lead to educational improvement. Studies done since the 1960s, have supplied plenty of evidence for one of the most important variables affecting academic attainment of those students, i.e., their level of language proficiency (Sharon, 1972; Mostiçyan, 1979; Douglas, 1986; Heikinheimo & Shute, 1986; Barber et al., 1987; Kinnell, 1990; Vinke & Jochems, 1993; Hawkey, 1982 cited in Tonkyn (1995); Tonkyn, 1995). That is why the questions of “What is language proficiency?” and “How is it measured?” will be posed and dealt with before going on to the rest of the study.

The examination of the recent literature in the field reveals that the nature and specification of the elements of language proficiency have not been determined and there continues to be a debate among academicians and practitioners about its definition (Oller & Damico, 1991). Hernandez-Chavez et al. (1978) argue that language proficiency consists of 64 separate language components, each of which is independently measurable. Oller (1978:143), on the other hand, claims that “there exists a global language proficiency factor which accounts for the bulk of the reliable

variance in a wide variety of language proficiency measures” (also see Oller & Perkins, 1978 and Oller, 1979). Clark (1972:5), however, looks at language proficiency from a different perspective. He defines it as the language learner’s ability “...to use language for **real-life** purposes without regard to the manner in which that competence was acquired.”

Considering the difficulty of defining language proficiency, it is conceivable that the development and use of proficiency tests would involve more complex steps than other types of language exams (Farhady, 1982). The difficulty stems from the fact that each language proficiency test should be based on a defensible model or definition of language proficiency (Bachman, 1990). That is why in their attempt to define and identify the ways of measuring language ability language testers devote much time and effort to define and identify the construct of language proficiency as well. According to Spolsky (1968:79), “fundamental to the preparation of valid tests of language proficiency is the theoretical question: What does it mean to know a language?” He argues that the answer to this interrogation is important since “which answer we accept has fundamental consequences for language testing, for how we go about measuring something is dependent on what it is that we think we are measuring” (Spolsky, 1985:181).

Hinofotis (1981), Spolsky (1989), and Brown (1996) have pointed out that language testing can be broken into periods, or trends, of development. Spolsky (1976) labeled them - 1) the Psychometric-Structuralist Approach; 2) the Psycholinguistic-Sociolinguistic Approach; 3) the Communicative Approach and 4) the General Proficiency Approach. Brown (1996:23), however, points out that he prefers to call

these trends “movements” because “they overlap chronologically and can be said to co-exist today in different parts of the world”.

The following section will examine the theoretical backgrounds of the language testing approaches and the types of tests which were developed as a result of these approaches.

1.1.1 The Psychometric-Structuralist Approach

The psychometric-structuralist approach assumes that the “knowledge of a language is best described, as the language itself, by describing its structures” (Spolsky, 1985:181). Proponents of this model see the language as “the sum of a variety of relatively characterizable parts or features that can be sampled, tabulated, and stated in some set of terms that correspond to greater or lesser proficiency in a language” (Gradman & Reed, 1997:199). According to Spolsky (1985:182), the structuralist approach is “most likely to be a competence model” since it prefers to deal with underlying knowledge, instead of attempting to show how the organization of the knowledge has direct results in performance.

The bridge between the structuralist language description and language testing was established by Lado’s (1961) theory of linguistic analysis. In his classical book on language testing Lado (1961) showed that the linguistic skills could be broken down, in accordance with contemporary structural linguistic theory, into their smallest components, permitting the tester to focus on precise areas of difficulty.

Language tests based on this approach are what Carroll (1961) has called “discrete point tests”, since they test knowledge of individual or discrete items selected from the structural description of the language (Spolsky, 1989). In this model the ideal assessment involves the evaluation of each of the domains of the structure and each of the skills of interest separately. Then, all the results are combined to form a total picture of language proficiency (Oller & Damico, 1991). Stated differently, an individual’s measure of language proficiency becomes a count of the number of items answered correctly. According to the proponents of this (psychometric-structuralist) approach, a test could not be valid if it attempted to measure more than one skill or domain at the same time (Lado, 1961).

Clear advantages of testing ‘discrete’ linguistic points are that they yield data which are easily quantifiable, as well as allowing a wide coverage. Discrete-point language tests typically use “objective” test formats. In other words, the examinees are required to select the appropriate option from a set of fixed choices. That is why these exams are efficient and have the usual reliability of marking associated with objectively scored tests (Weir, 1990).

According to Hinofotis (1981) and Brown (1996), the psychometric-structuralist approach is an important step in the history of language testing since for the first time language test development started using scientific principles. In addition, structuralist tests are still very much in evidence around the world. This approach led to the development of standardized tests like the Test of English as a Foreign Language (TOEFL - first introduced in 1963), the Michigan Test of English Language Proficiency: Form A (MTLP - University of Michigan, 1961), the

Modern Language Association Foreign Language Proficiency Tests for Teachers and Advanced Students (Educational Testing Service, 1968) and the Comprehensive English Language Test for Speakers of English as a Second language (Harris & Palmer, 1970).

However, structuralist tests have been harshly criticized due to their three important features (Oller and Damico, 1991). Firstly, they try to limit language testing to a single skill without involving or including another (e.g., intend to measure writing without reading). Chaplen (1970:27)* criticized isolated skill tests arguing that: "It seems unlikely that measurement of the component skills most commonly isolated can provide either singly or in aggregate, a satisfactory measure of the gestalt." This view was backed up by Savignon (1972), who found that grammatical competence was not by itself a good predictor of communicative skills.

Secondly, discrete point tests intend to limit language testing to a single domain without including other domains (e.g., try to measure vocabulary without including phonology). However, Spolsky (1968) argued that perhaps instead of attempting to establish a person's knowledge of language in terms of percentage of mastery of morphology and lexicon, it would be better to try to test that person's ability to perform in a specific socio-linguistic setting.

Finally, discrete point language tests try to measure language in isolation. That is, they do not take into consideration such factors like social context or human experience.

Oller (1979:212) expressed his criticism in relation to this feature of discrete point language tests in the following way:

Discrete point analysis necessarily breaks the elements of language apart and tries to teach them (or test) them separately with little or no attention to the way those elements interact in a large context of communication. What makes it ineffective as a basis for teaching or testing languages is that crucial properties of language are lost when its elements are separated. The fact is that in any system where the parts interact to produce properties and qualities that do not exist in the part separately, *the whole is greater than the sum of its parts*.... organizational constraints themselves become crucial properties of the system which simply cannot be found in the parts separately.

The difficulty of limiting language testing to a single skill and domain, or the impossibility to extract the language from its natural social context led testing experts to look for other ways to measure the complex variable named “language”.

In response to the feeling that discrete point tests were not sufficient indicators of language proficiency, the testing pendulum on the whole swung in favor of global tests in the 1970s. Spolsky (1978) termed this new approach to language testing “the psycholinguistic-sociolinguistic approach”. According to Weir (1990:3) this was “an approach to measure that was in many ways contrary to the allegedly atomistic assumptions of the discrete point tests” (also see Davies, 1978a and 1978b).

* Cited in Weir (1990).

1.1.2 The Psycholinguistic-Sociolinguistic Approach

The psycholinguistic-sociolinguistic approach has its roots in the argument that language is creative. More precisely, language professionals began to believe that language is more than the sum of the discrete parts being tested during the psychometric-structuralist movement (Brown, 1996). In its current form, this approach is embodied variously in the notional-functional curriculum, the communicative competence model, and the interest in teaching and testing pragmatics.

Read (1981:x) succinctly described the psycholinguistic-sociolinguistic era:

From a psycholinguistic perspective, language came to be seen as less of a well-defined taxonomic structure and more of a dynamic, creative, functional system. It was recognized that natural language contains a considerable amount of redundancy, so that it is difficult to show that any single linguistic unit is indispensable for communication... . The sociolinguistic contribution centers on the concept of communicative competence, which represents a broadening of Chomsky's notion of competence to cover not only knowledge of rules for forming grammatical sentences but also rules for using those sentences appropriately with different context....

Beginning with the work of sociolinguists like Hymes (1967), it was felt that the development of communicative competence depended on more than a simple grammatical control of the language. Hymes (1972) proposed a two-dimensional communicative competence model, comprising of "linguistic" and "sociolinguistic" elements. That is, communicative competence, according to Hymes (1972), included

the ability to use the language appropriate for different situations, as well as having the knowledge, which underlies this usage.

Depending on Hymes's (1972) and other current linguists' works, testing experts started to scrutinize the language from another perspective. They identified and listed the various possible functions of language, including all the notions that could be expressed in it depending on the social context. This angle gave testers an opportunity to handle with variability more easily than structuralists did. Actually, the functional movement was the approach to language that led to the notion of teaching language for specific purposes, i.e., to teaching selected register.

As it was mentioned previously, according to the proponents of the psycholinguistic-sociolinguistic approach "language processing or use entails the simultaneous engagement of more than one language component (e.g., vocabulary, grammar, gesture) and skill (e.g., listening, speaking)" at a time (Del Vecchio & Guerrero, 1995:6). Moreover, according to them language proficiency should be assessed in a fairly rich context of discourse since it was inappropriate to think of language in isolation (Carroll, 1961; Oller, 1979). Due to these concerns, language testing experts designed and started to employ the so called "integrative tests". Tests that integrate a large number of different discrete items by calling on the subject to perform some function or task using the target language. For example, an integrative task may ask the test-taker to listen to a poem, then to read a second poem, and then to write a story on the topics of those two poems, and at the end to retell his/her story. Examples of such tests are the BUEPT (Boğaziçi University, 1983), the Certificate of Proficiency in English (CPE) and the writing part of the TOEFL.

Typical test formats of this movement are the cloze test and dictation. Brindley (1986) argues that through these tests examinees have to mobilize their linguistic and extra-linguistic knowledge to reconstitute the meaning of a text. Oller (1976), on the other hand, maintains that tests such as cloze and dictation require 'performance' under real life constraints, i.e., they guide to aptitude and potential for communication, even if they do not test communication itself. Moreover, according to Oller (1979), they are practicable to administer, economical to set and mark, and have respectable reliability figures associated with them.

Oller (1979) also stated that his integrative tests were representing total language proficiency better than any other single test or combination of tests. There were empirical and theoretical arguments presented for this claim. Oller and his colleagues, for example, used factor analyses to uncover the relation between the cloze tests and dictations, and other types of tests. Researchers were struck by the power and importance of a common, first factor that Oller labeled "unitary language competence".

Studies conducted by a number of testing experts, however, revealed that there were some serious problems related with cloze tests and dictation, and that these problems were affecting both the validity and the reliability of these integrative exams (Alderson, 1978; Morrow, 1979; Carroll, 1980; Hughes & Porter, 1983; Weir, 1990).

Work by Alderson (1978) raised serious questions about validity and reliability of cloze tests as testing devices. Experiments demonstrated that the scores taken on these tests were affected by altering the point where the deletions were started from,

or by using the different n/th rate deletion (even when the same passage was used). Depending on these results, Alderson concluded that there was no such a thing as “the cloze test”.

Hughes and Porter (1983) criticized the statistical analysis approach employed by Oller and his colleagues. They (Hughes & Porter, 1983) emphasized that the factor analysis tends to exaggerate the size of the first factor when another hypothetical model is lacking. Oller (1984) has acknowledged this criticism and is now much more hesitant in his claims.

Another piece of evidence against the cloze tests comes from Morrow (1979). He states that they are tests of underlying ability (competence) and they do not give examinees an opportunity to show their actual performance. Carroll (1980:9) backs up Morrow's conclusion. He argues that cloze test “is still essentially usage based. The task does not represent genuine interactive communication and, is therefore, only an indirect index of potential efficiency in coping with day-to-day communicative tasks.”

The artificiality of the type of information supplied by the “discrete point” and “integrative” tests resulted in a demand for tests which would require tests takers to produce ‘real’ language. Hinofotis (1981) in her discussion of the trends of 1980s, suggests that the influence of the communicative revolution has added new elements to language testing. She points out that the new models of communicative competence as Canale and Swain (1980), Canale (1983), and Bachman's (1990) gave rise to a new trend named “the communicative movement”.

1.1.3 The Communicative Approach

During the communicative teaching era of the 1970s, which was critical of traditional non-communicative tests, a number of language testers called for the development of tests that would provide a better reflection of the real-life language use (Clark, 1976; Jones, 1977; Morrow, 1977). The claim that the language produced on non-communicative tests was artificial gave rise to a third trend in language testing, i.e., the pragmatic or task-based tests. That is, tests that require the students to use the language to perform a task. Clark (1976), for example, referred to task-based tests as those in which real-life tasks with real constraints, inputs and outputs were replicated as far as possible. His example of a task-based test of oral language was one involving a test setting where both the examinee and the testers were engaged in a communicative dialogue and performed in actual communicative situations. Morrow (1977), on the other hand, proposed offering tests takers the opportunity for spontaneous operation of the language in authentic settings and activities that simulated performances they would have to engage in outside the test situation.

Testers operating within the communicative paradigm soon realized that it was naive to assume that one could develop valid tests of communicative language ability without reference to the construct which one was attempting to measure (Weir, 1990).

The first theoretical model intending to define the construct named “communicative competence” was that of Canale and Swain (1980). However, it is really important

to place their framework within the time that it was introduced before explaining the model itself.

The development of the theory of communicative competence was an indirect result of the Chomskyan revolution in linguistics, which as Spolsky (1989) notes, had an indirect effect on theories of second language learning. In mid 1960s, Chomsky (1965) distinguished between linguistic “competence” and “performance”. He pointed out that there was a fundamental difference between “the speaker/hearer’s knowledge of his/her language” (competence) and his/her actual use of language in concrete situations (performance). Taylor (1988) argued that by the term “competence” Chomsky referred exclusively to knowledge, not including the notion of capacity or ability. Thus, the Chomskyan notion of linguistic competence refers to “the underlying knowledge of an idealized native speaker of a language that enables such person to distinguish grammatical from ungrammatical sentences of the language” (Spolsky, 1989:51). Spolsky (1989) notes that all Chomsky’s examples make it clear that he is using the term “competence” quite differently from its normal use, and that it is a mistake to confuse the Chomskyan notion of competence with the general notion of ability. In his theoretical work, Chomsky chose to limit his attention to the rules of language needed for generative grammar, restricted to sentence-length utterances and exclusive variation.

In the term “performance”, Chomsky referred to a large array of factors such as limitation of memory, which explains the constraint on the length of sentences that are grammatically infinite. For Chomsky performance implies actual, real-life

performance. Schachter (1990) points out that it should not be mixed with the realization of ability or potential.

Chomsky's somewhat limiting definition of the scope of linguistic theory left the way open for Hymes (1972) to propose the complementary notion of communicative competence, in which the focus is not on a well formed sentence but on one which is appropriately used in a specific context. Hymes (1972) differentiated between linguistic and communicative competence, and between linguistic and communicative performance. Thus the term "performance", according to Hymes refers to actual use and actual events and ability for use.

Discussion presented above reveals that an agreement on what components should be included in a model of communicative language ability is by no means unanimous (Courchene & de Bagheera, 1985:49). Canale and Swain (1980), and Canale's (1983) model is even more elaborated than the previous ones. It involves recognizing four distinct but related competencies:

1. *Grammatical Competence*: mastery of language code (e.g., vocabulary and rules formation, sentence formation, literal meaning, pronunciation, and spelling).
2. *Discourse Competence*: mastery of how to combine meanings and forms to achieve unified text in different genres (e.g., causal conversation, an argumentative essay, or a business letter) by using both cohesion devices to relate forms and coherence principles to organize meanings.

3. *Sociolinguistic Competence*: mastery of appropriate use (production and comprehension) of language in different sociolinguistic contexts, with emphasis on appropriateness of meaning (e.g., topics, attitudes, functions), and form (register, formulaic expressions).
4. *Strategic Competence*: mastery of verbal and nonverbal strategies both to compensate for breakdowns in communication due to insufficient competence or to performance limitations, and to enhance the rhetorical effect of language (Duran et al., 1987:9).

The examination of the model proposed by Canale and Swain reveals how complex is the construct (i.e., language proficiency) that is tried to be measured. The model shows that besides the linguistic competence emphasized by structuralist description there are three other full sets of rule-governed language systems that should be taken into consideration while preparing a language test.

The unique appeal of the Canale and Swain's (1980) framework was that it did not adopt Chomsky's (1965) distinction between competence and performance, but rather related more to Hymes's (1972) notion of communicative competence. Their argument was that linguistic competence should be viewed as part of communicative competence since rules of grammar are meaningless without roles of use.

Recently, another theory of communicative competence was developed by Bachman (1990). He named his model "Communicative Language Ability Model". The framework proposed by Bachman (1990) gathers language ability around organizational and pragmatic competencies. Organizational competence consists of

grammatical and textual competencies, while pragmatic competence consists of illocutionary and sociolinguistic competencies. According to Bachman (1990), the model consists of both knowledge (what he also calls competence) and the capacity for implementing, or executing that competence, in appropriate contextualized, communicative, language.

Weir (1990) states that language models as the ones presented above provide a potentially useful framework for the design of language tests.

Morrow (1979) and Canale and Swain (1980) argue that the aim of pragmatic language testing is three focal. That is, those tests besides being interested in what the learner knows about the form of the language, about how to use it appropriately in context of use, must also deal with the extent to which the learner is actually able to demonstrate this knowledge in a meaningful communicative situation (performance), i.e., what he can do with the language.

Rea (1978:4), however, argues that the main goal of pragmatic tests is to measure the tester's "ability to communicate with ease and effect in specified sociolinguistic setting." According to Oller (1979) this is the most striking difference between integrative and pragmatic tests. He (1979:38) points out that "pragmatic tests are always integrative", but they differ from integrative tests in that they require the test-takers to engage in different tasks only if the contextual and temporal conditions that generally characterize this activity are fulfilled. For example, in pragmatic listening tests lectures are usually delivered "live", instead of using a tape-recorder. This is due to the fact that in real life lectures are nearly almost delivered 'live'.

The capacity to use language communicatively thus involves both competence and demonstration of the ability to use this competence (Widdowson, 1983; Bachman, 1990). It is believed that the performance tasks candidates are faced with in communicative tests should be “representative of the types of tasks they might encounter in their own real-life situation and should correspond to normal language use where an integration of communicative skills is required with little time to reflect on, or monitor language input and output” (Weir, 1990:9). There is a clear shift in emphasis from the linguistic to communicative dimension. Stated differently, now the sociolinguistic and strategic competencies in the Canale and Swain’s model are emphasized.

The emphasis is no longer on linguistic accuracy but on the ability to function effectively through language in particular context of situation. It is therefore essential to be as precise as possible about the skills and performance conditions for any tests which claim to assess communicative language ability. Skehan (1988) argues that test constructors must closely identify those skills and performance conditions that are the most important components of language use in particular context. This, however, raises the problem of generalizability. Morrow (1977:53) was aware of this problem related with pragmatic tests. He set it out like this:

The very essence of a communicative approach is to establish particular situation with particular features of context, etc., in order to test the candidate’s ability to use language appropriate in terms of a particular specification. While it is hoped that the procedures discussed will indeed be revealing in those terms, they cannot strictly speaking reveal anything of the candidate’s ability to produce language which is appropriate to a situation different in even one respect from that established.

Alderson (Alderson & Hughes, 1981) also accepted that in order to design a pragmatic test one needed to define what was that learners were expected to do with language in a specific situation. Nevertheless, he recognized that by specifying performance in this way "one might end up describing an impossible variety of situations, which one can encompass for testing purposes" (ibid., 59).

As seen from the reviews above each of the test movements emphasizes a different part of the construct called "language proficiency". While discrete point tests have been interested only in measuring language knowledge, integrative tests widened this scope and included discourse competence besides language knowledge. Task-based (pragmatic) tests, however, emphasize the role of sociolinguistic and strategic competencies, as defined by Canale and Swain (1980). However, none of these approaches answers the crucial question of how the parts of the communicative competence are related to each other.

Recently, the legitimacy of separate skills testing is being questioned and indeed the more innovatory testing of skills through an integrated story-line of procedure (see Low, 1986) is gaining favor. Spolsky (1985) states that he is willing to treat language ability as a General Proficiency. He argues that showing that "linguistic and communicative competence are divisible does not necessarily rule out the claim that there is a core of common knowledge of language underlying the specific abilities of the speaker" (ibid., 185).

1.1.4 The General Proficiency Approach

The general proficiency claim, also referred to as the unitary competence hypothesis, was originally derived from Carroll's (1961) idea of integrative language tests. The argument was presented by Carroll as follows:

The high correlation obtained between various sections of TOEFL and other general test of English suggests that in fact we might be dealing with a single factor, English proficiency...

A decade later the theoretical argument for the unity of the language proficiency was presented by Spolsky (1971). He focused attention on the link between the creative aspect of language knowledge and the ability to operate in conditions of reduced redundancy (see Miller and Isard (1963) work in information theory).

As a result of these Oller (1976, 1979, 1981) promoted his notion of language as a unitary factor rather than as a divisible construct. He claimed that the learner in the course of language acquisition was developing a central, comprehensive and integrative linguistic competence which was going to serve all forms of language use. Later, however, Oller (1983:36) accepted that the strong form of his hypothesis was wrong and he explained his current position with the following words:

...the holistic, global aspects of language use, or other cognitive performances, do not exclude particulate, analytic and discrete elements. Indeed it seems that in an adequate theory the holistic elements must depend on the interaction of the relatively analytic components. Hence the idea that global and particulate models were incompatible must have been wrong...It would seem that both views are needed and that they can complement each other rather than contradict one another.

The general proficiency claim, is based on the notion that "there is some fundamentally indivisible (even is technically analyzable) body of knowledge varying in size from individual to individual such that you can rank individuals on the extent of their knowledge" (Spolsky, 1989:71). Spolsky (1989:72) claims that even if it is not possible to specify a minimum of structural knowledge or communicative competence of a particular individual, some general claims about the ranking of this individual relative to him or herself at other times or relative to other people can be made. Statements like "X knows more of this language now than last month" or "X knows more than Y" can be uttered in order to express the subject's control of the language on a single gradient. The choice of the term "proficiency", the emphasis on discovering it through test performance and the requirements of naturalness imply that this is a model of performance rather than competence (Spolsky, 1985).

In testing, the claim for general proficiency underlies the discussion of the scores taken either on batteries of tests of various kinds or on the privileged test methods like cloze and dictation.

The discussion so far has shown that there are both in general theory and in language testing theory four interrelated but not overlapping approaches to describing and measuring knowledge of second language: the one structural, the second functional, the third communicative and the fourth general. Spolsky (1985) argues that anybody who knows a second language may be assumed to have all four kinds of knowledge. According to Spolsky (1989), the relations between the kinds of language knowledge are hierarchical. That is, the more structural items the

learners control, the more functional ability they have; the more functional ability they possess, the better they use the language in real-life; and the better they use the language in real-life, the greater the level of their general proficiency is (Spolsky, 1989:79).

However, the problem of testing language proficiency does not end after designing the test. As Tonkyn (1995:39) points out “whatever the nature of the test, it might be said that the proof of the pudding should be in the eating”. There is a need to periodically evaluate tests that are already operational. For example, initial validity study may not be sufficient to guarantee the validity of a test throughout its life time (Frisbie, 1982).

Besides that, as Fredericksen and Collins (1989) pointed out since the “educational system adjusts its curricular and instructional practices and students adjust their learning strategies and goals to maximize the score on the test”, the validation of the test employed in the educational system becomes imperative. Then it is necessary to search for answers to the following questions: 1) What is validity? 2) Why should a language proficiency test be valid?

In the following section the term validity will be presented and discussed.

1.2 Validity

Test experts have always been interested in “test validity” (Gray, 1997:1). It is “the most important consideration in test evaluation” (American Educational Research Association, American Psychological Association, and National Council on Measurement and Education, 1985:9). According to Hubley and Zumbo (1996:207) “of all the concepts in testing and measurement validity is the most basic and far-reaching; for without validity, a test, measure or observation and any inferences made from it are meaningless”. Thus it is important to give the definition of validity before going on to the discussion of its different types (e.g., construct, content, predictive).

One of the earliest definitions of validity comes from Cronbach (1949:48) in his classical book named *Essentials of Psychological Testing*. He points out that “a test is valid to the degree that we know what it measures or predicts”. A more widely cited definition is composed by Anastasi (1954:120): “Validity is what the test measures and how well it does so”. However, the definition that is considered in the current study is the one by Messick (1995:741), which states that “Validity is an overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other models of assessment.” According to Messick (1995), validity is not a property of the test but rather of the meaning of the test scores. In addition, these scores are functions not only of the test items and test conditions, but also of the persons responding and the context of the assessment.

Stated differently, what needs to be valid is the meaning of the scores; as well as any implications for action that this meaning entails (Cronbach, 1971).

In the early 1950s, there was a proliferation of different concepts and delineation of validity. Some of the validity types proposed include Guilford's (1946) factorial and practical validities; Mosier's (1947) face validity, Gulliksen's (1950) intrinsic validity, Anastasi's (1954) face, content, factorial, and empirical validities and American Psychological Association's (APA) (1954) content, predictive, concurrent and construct validities. However, for the purpose of this study only predictive and face validities will be presented and discussed in detail*.

1.2.1 Predictive Validity

Predictive validity that was defined as parallelism between the test scores and the criterion performance (Angoff, 1988) was often considered to be the most valuable type of validity (Hubley & Zumbo, 1996). Bachman (1990:253) also pointed out that prediction is an important and justifiable use of language tests. According to him, the evidence that indicates a relationship between test performance and the behavior that is to be predicted provides support for the validity of this use of test results.

Predictive validity of various language proficiency tests is of particular interest to the institution, as results received in the exams are important part of the criteria drawn on to assess students' suitability for admission to its various academic programs

* For the detailed description of the other types of validity, see Cronbach (1984), Heaton (1988), Hughes (1989), Messick (1989), Bachman (1990), and Hubley and Zumbo (1996).

(Paltridge, 1992:261). Predictive validity becomes even more important when language proficiency tests are evaluated, since they are expected to “look forwards rather than backwards” (Alderson, 1984:33).

In the following section studies which have been conducted to explore the predictive validity of different English proficiency tests will be discussed. Since studies on the relationship between English proficiency tests and academic achievement have indicated a wide range of findings and implications, they are not discussed according to such criteria as correlation coefficients, but according to the conclusions reached by researchers, in the following categories: 1) positive, 2) negative, or 3) mixed.

1.2.1.1 Positive Conclusions

A number of studies evaluating the relationship between the scores of English language proficiency tests and academic success found a positive relation between these two criteria (Burgess & Greis, 1970; Heil & Aleamoni, 1974; Ayers & Peters, 1977; Baldauf & Dawson, 1980; Freidenberg & Curry, 1981; Rea, 1984; Ho & Spinks, 1985; Douglas, 1986; Dedo, 1990; Paltridge, 1992; Wen & Johnson, 1997).

Some of the studies that ended up with positive conclusions investigated the relationship between the English proficiency test scores and the freshman year academic success of EFL/ESL students.

Heil and Aleamoni (1974) conducted a study aiming to evaluate the predictive validity of TOEFL. In the study, TOEFL scores of foreign language students in American Universities were correlated separately with their first and second semester freshman year GPAs. Results indicated that TOEFL predicts second semester GPAs ($GPA\ r = .336$) better than first semester GPAs ($GPA\ r = .270$). Considering the results Heil and Aleamoni (1974) concluded that TOEFL was a useful predictor of the academic success of nonnative English speakers in American Universities.

Two other important studies were those of Ho and Spinks (1985), and Dedo (1990). Methodologies used there were different from the one employed in the study discussed above. Instead of correlating GPAs only with the scores of proficiency tests, researchers set to determine which one among different factors (such as educational, social and writing background, English language proficiency tests scores, learning strategies, and motivation) best predict the freshman year academic performance of university students. Multiple regression analysis revealed that the scores on the English tests had the most predictive value. Considering the results, Ho and Spinks (1985:258) concluded as follows: "it is quite certain that students who are deficient in English would be handicapped in their learning at the University". Dedo's (1990) conclusions were not different.

Wen and Johnson (1997) also found that the level of English proficiency, as measured by the English matriculation examination in China, was one of the best predictors of the freshman year academic achievement of 242 English major university students in Nanjing and Shanghai.

Douglas (1986) carried out an interesting research with 75 Sudan students at the University of Khartoum. The aim of the study was to investigate whether or not English proficiency tests employed by the university successfully predict the overall freshman year academic attainment of ESL subjects who had graduated from vernacular secondary schools. The results revealed that all three of the English proficiency measures significantly correlated with students' freshman year cumulative GPAs.

The second group of studies set out to investigate whether the different English language proficiency tests (e.g., MTELP, TOEFL) were valid predictors of the general academic attainment in an ESL/EFL tertiary educational setting or not. (Burgess & Greis, 1970; Baldauf & Dawson, 1980; Freidenberg & Curry, 1981).

Burgess and Greis (1970) carried out a study with the aim of correlating the TOEFL scores and the overall college GPAs of 17 ESL students. To attain valid results, researchers conducted a two-step statistical analysis. TOEFL scores were first correlated with the overall GPAs of students, and then with their weighted GPAs (i.e., the grades in such courses, as music, art and maths that require little English, were deleted). The findings of the study indicated that TOEFL scores did predict the future academic success of subjects. Both of the correlations were significant, with the one between the TOEFL scores and the weighted GPA being slightly higher than the one between the TOEFL scores and the overall GPA (TOEFL with overall GPA $r = .53$; TOEFL with weighted GPA $r = .56$).

Another research aiming to evaluate the relationship between the scores of English language proficiency tests and the overall academic success was conducted by Baldauf & Dawson (1980). The survey took place in a teacher-training college in Papua New Guinea. The subjects were students who entered the university between 1978 and 1980. The correlations between the MTELP results and the overall GPAs of the subjects varied from .33 to .74. for the different groups of students.

A year later a study intended as a replication of the research by Baldauf and Dawson (1980), was carried out with forty-two Cuban American students enrolled in a bilingual teacher education program at Florida International University (Freidenberg & Curry, 1981). Here again a significant relationship was found between the scores of MTELP and the overall college GPAs of subjects ($r = .41$). The common conclusion drawn was that the MTELP had a significant predictive validity as a measure of the overall academic achievement of university students. The researchers also pointed out that their studies proved that English language competence was important for "ESL student teachers regardless of their intended teaching specialty when English is the language of instruction" (Baldauf & Dawson, 1980:1204).

One of the recent studies in the area was conducted by Patridge (1992). He developed an English for Academic Purposes (EAP) placement test, which was aimed to reflect the integrated approach of language use. Paltridge reported that the test that was administered both in Japan and New Zealand appeared to be successful in distinguishing between undergraduate students who were proficient and those who needed remedial language work. Patridge (1992:243) argues that there is a need for the development of proficiency tests reflecting an integrated approach. He states that

although the “practices in areas of second and foreign language teaching and learning have changed dramatically and have become communication-based and learner-centered”, language testing is still structure-based.

Differently from the studies discussed above, Ayers and Peters (1977) were the two experts who investigated the relationship between the English proficiency test scores and the academic performance of graduate EFL students. The subjects of the study were 50 Asian students who were completing masters programs in engineering, chemistry, or mathematics. Their overall GPAs were first correlated with their TOEFL scores, next, with the scores obtained on the Verbal Ability section of the Graduate Record Examination (GRE-V), and finally with the average of scores obtained on TOEFL and GRE-V. It was reported that the TOEFL was a better predictor in determining the success of the Asian students in the study than GRE-V (GPA with TOEFL $r = .40$; GPA with GRE-V $r = .22$). However, the highest correlation found was between GPAs and the combination of scores from TOEFL and GRE-V ($r = .55$).

1.2.1.2 Negative Conclusions

A large number of studies on the relationship between English language proficiency tests and academic success have led researchers to negative conclusions (Mulligan, 1966; Sugimoto, 1966; Hwang & Dizney, 1970; Sharon, 1972; Gue & Holdaway, 1973; Andalib, 1975; Shay, 1975; Wilcox, 1975; Odunze, 1980; Light et al., 1987; Vinke & Jochems, 1993; Morrison & Morrison, 1995; Yan, 1995). Moreover,

Tonkyn (1995) concluded that none of the major standardized language tests used in various British universities to predict the future success of students – e.g., Davies English Proficiency Test Battery (DEPTB), English Proficiency Test Battery (EPTB), English Language Battery (ELBA), International English Language Testing Service Test (IELTS) – performed particularly well.

Hwang and Dizney (1970) in their study of 63 Chinese university students at the University of Oregon found no significant correlation between TOEFL scores and the freshman year first-term GPAs ($r = .19$). The results of the study were very important for learners were majoring in subject areas like education, social sciences, and architecture that require a high level of English proficiency.

Research aiming to evaluate the predictive validity of TOEFL was conducted by Gue and Holdaway (1973). In this study, however, TOEFL scores were correlated with the overall freshman year GPAs of 123 Thai education majors at the University of Alberta. Scores on TOEFL sub-tests (i.e., listening, reading, vocabulary, English structure, and writing), total summer and fall TOEFL scores, and final GPAs were intercorrelated. The results indicated that neither the total TOEFL scores nor the sub-test scores were good predictors of the freshman year academic achievement as measured by GPAs. The authors argued that a possible reason for the low correlation between the overall TOEFL scores and GPAs might be due to a mismatch between the testing methods. It was reported that in contrast to TOEFL, which is a multiple choice test, a considerable weight was apparently placed upon subjective written examinations at the University of Alberta. Another conclusion reached by the researchers concerned the administration of the listening comprehension sub-test.

Gue and Holdaway (1973:101) claimed that the “thorny problem of the accent of the reader of the lecture” and the nature of the simulated talk were factors that reduced the reliability and validity of the listening comprehension part of the test. According to them, a different factor that affected the strength of the predictive validity of the test was the organizational climate in a test situation.

Odunze (1982) carried out a similar study with 118 Nigerian students in four Missouri universities. He also did not find a significant correlation ($r = .259$) between TOEFL scores and “the first year school work” of his subjects. However, he proposed an explanation different from Gue and Holdaway’s (1973) for the lack of the relationship between the two variables. Odunze argued that TOEFL is not an appropriate tool to measure the level of English language proficiency of Nigerian students. He concluded that using the TOEFL as a major determinant of international students’ admission to various academic programs in colleges and universities, and as a predictor of international students’ subsequent academic success should be reconsidered. Trainor (1985) backed up Odunze’s recommendation in his study. After examining the structure and the content of TOEFL, he concluded that the exam “appears to suffer some very serious defects” such as, encouraging guessing, being culture biased, failing to test what it purports to test, and not measuring all four skills.

Later studies carried out by Light et al. (1987), and Yan (1995) had two goals: (1) to find the extent to which TOEFL scores predict the freshman year first semester academic success for international graduate students at the State University of New York at Albany (SUNYA) and Mississippi University (MSU) respectively; and (2) to

determine the relationship between the institutional admissions criterion of a TOEFL score of 550 and students' GPAs. The correlations between the TOEFL scores and the freshman year first semester GPAs of subjects from different departments were calculated. However, all of the obtained correlations were too low to have any practical significance. Conclusions drawn by researchers (Light et al., 1987; Yan, 1995) can be summarized with Light et al.'s (1987:255) sentence that "knowing how a student scored on TOEFL will tell us practically nothing we need to know to predict the student's academic performance."

As an answer to the second research question, researchers stated that the study did not provide empirical rationale for the TOEFL cut off point. Yan (1995) pointed out that no significant difference was found between the freshman year first semester GPAs of the international students whose TOEFL scores were at or above 550 and those whose scores were below 550. Light et al. (1987), on the other hand, reported that students with lower TOEFL scores (between 400 and 549) performed academically better (i.e., had higher GPAs) than those with higher TOEFL scores (between 550-569). Another salient finding of Light et al.'s (1987) study was that the relationship between TOEFL scores and GPAs of the social science group was significantly higher than the relationship between TOEFL scores and GPAs of science group students.

Another important study examining the predictive validity and the cut off score of TOEFL was conducted by Vinke and Jochems (1993). The subjects of the study were 90 Indonesian engineering students. This time, TOEFL scores were correlated with the average of the passing grades of seven written examinations. The TOEFL

cut off score was set as 420. The study ended up with a correlation coefficient of .51 between the two TOEFL scores and academic attainment. Nevertheless, Vinke and Jochems concluded that not the statistical, but the real significance of the test is important, i.e., percentage of variance (26%). According to them, the low percentage of variance is an indication of the fact that "language tests such as TOEFL do not measure the language skills that foreign students need in order to be academically successful at an English-medium university or college" (Vinke & Jochems, 1993:282). Another finding of the study concerned the TOEFL cut off points. Researchers reported that it did not seem to "make a difference whether students have TOEFL scores of, for example 460 or 520". No significant difference was found in the academic performance of students at both ends of the continuum (460 vs. 520). These results led to the formulation of the following hypothesis:

There is a range of TOEFL scores within which a better command of English increases the chance of academic success to a certain extent and within which a limited lack of English proficiency can be off set by greater student effort or greater academic abilities. The limits of this range still need to be established (Vinke & Jochems, 1993:282).

Another group of studies aimed to examine whether GRE-V, either in combination with TOEFL or alone, was a valid predictor of the overall graduate GPA of EFL students or not.

Sharon (1972) attempted to investigate whether the TOEFL added to the predictive value of GRE-V. He scrutinized the relationship between the overall graduate GPAs of 975 foreign graduate students from 24 schools in the USA, and their TOEFL and GRE-V scores. The research manifested that not only TOEFL failed to predict ($r =$

.26) the future academic success of subjects but also it added nearly nothing to the predictive feature of GRE-V ($r = .27$ when TOEFL and GRE-V scores were combined).

A study done by Shay in 1975 also aimed to examine how TOEFL and GRE-V's results correlate with the graduate academic success. Here again none of the correlations obtained were statistically significant (r s varied from .08 to .12 between TOEFL and freshman year GPAs; and r s varied from .06 to .09 between GRE-V and freshman year GPAs). Similarly to Sharon (1972), Shay also reported that these two tests fail to predict academic performance of graduate EFL students.

Recently, Morrison and Morrison (1995) did an analytical review of 22 studies examining the relationship between the performance on the GRE-V and the graduate GPAs of foreign university students conducted between 1955 and 1992. They pointed out that the inability of GRE-V to predict the academic success of foreign university students was the most striking finding of their study.

1.2.1.3 Mixed Conclusions

Another group of studies of the relationship between English language proficiency and academic success led the researchers to mixed conclusions (Abadzi, 1975; Bostic, 1981; Mestre, 1981; Slark & Bateman, 1982; Stover, 1982; Zeidner, 1986, 1987; Robinson & Ross, 1996).

Abadzi (1975) carried out a study with the aim of correlating the TOEFL scores with the first and second semester freshman year GPAs of 70 foreign students who entered the University of Alabama in 1974-75. A strong relationship was found between the TOEFL scores and the first semester GPAs of students ($r = .43$), however, no correlation was found between the second semester GPAs and the English knowledge as measured by TOEFL.

A later study was carried out by Stover (1982). Here TOEFL scores and preuniversity English program (prep) grades were correlated with the first semester GPAs of 159 undergraduate and graduate students at the University of Arizona. TOEFL appeared to be a good predictor of the first semester academic attainment of both graduate and undergraduate subjects. All subjects in the study who scored 500 and above on TOEFL reached "an acceptable level" of achievement at the end of the first semester. Nevertheless, prep grades could only predict the freshman year first semester GPAs of undergraduate students ($r = .21$). They were not found to be significantly related to the success of graduate students.

Bostic (1981) conducted a study intending to correlate the overall freshman year GPAs of 154 foreign students at Oklahoma College, attending both "language oriented fields" and "scientifically oriented fields", with their TOEFL scores. The study showed that there was a significant, but not strong, positive relationship between the overall freshman year GPAs and TOEFL results ($r = .169$). The salient finding of this study was the fact that TOEFL scores significantly correlated with the overall freshman year GPAs of the students who studied in "scientifically-oriented fields" as compared to the ones who studied in "language oriented fields".

Two studies carried out by Zeidner (1986, 1987) explored the predictive validity of College Admission English Language Aptitude Test (ELAT). The researcher correlated first year cumulative GPAs of Jewish and Arab students at the university in Northern Israel with ELAT test scores. The relationship between the associated criteria was found to be significant. However, the exam seemed to mispredict the success of both groups. That is, it overestimated the academic attainment of Arab students and underestimated the success of Jews. In spite of this finding, Zeidner suggested that in principle, ELAT might be used fairly for predicting the academic achievement of varying cultural groups, especially Arabs.

In another study, Mestre (1981) intended to examine the relationship between the results of English language proficiency tests and the overall college GPAs of bilingual Hispanic and monolingual American students. The level of English language proficiency of students was measured by Manuel's Reading Test (1962) and the Verbal Ability Section of the Scholastic Aptitude Test (SAT-V). Results of the study revealed that Manuel's Reading Test successfully predicted the future academic achievement of both groups of students. However, SAT-V results were significantly related only with the monolingual group's GPAs. This finding led Metsre to question whether or not it is appropriate to use SAT-V as a basis for admission decisions for bilingual students.

Finally, Robinson and Ross (1996) conducted a study comparing the predictive validity of two English for Academic Purposes tests. The first of the tests was a Task-based Performance Test (TPT) designed by researchers. The second test was a traditional skill-focused English Language Institutional Placement Test (ELAT).

Academic success was defined as a pass/fail mark on the direct performance-reference task. The TPT was found to predict subjects' success better than the ELAT. According to Robinson and Ross (1996:467) this was due to the fact that skill focused tests did not provide sufficient grounds for predicting the actual academic skills of students. According to them, these kinds of tests only provide an indication of the current knowledge of English language.

1.2.2 Face Validity

Face validity pertains to whether or not the test "looks valid" to the examinees who take it. Although some measurement professionals have never seriously considered "face validity" (Mosier, 1947; Cattell, 1964; Davies, 1977; Stevenson, 1981, 1982, 1985; Cronbach, 1984; Bachman, 1990), others have claimed and empirically proved that it is an important feature of tests (Frisbie, 1982; Low, 1982, 1985; Shohamy, 1985; Mendelsohn, 1989; Anastasi & Urbina, 1997; Fulcher, 1997). Another research that provided a piece of crucial evidence for the importance of face validity was conducted by Jonson and Plake (1998). Researchers (Johnson & Plake, 1998) examined the preview descriptions of validity expressed in the *Mental Measurement Yearbook Test Reviews* for the last 50 years. They reported that the requirement of "face validity" appeared in every test review that was examined.

Low (1985:156) argues that testees are the group from whom data related with the face validity of a test should be gathered. He states that the test-taker is the only

person with any direct knowledge of “whether the form of presentation in which the test content is couched is such that the desired linguistic behavior is not produced” (ibid., 156). Low also adds that testees can often feel that this is “not how they read articles in real life” (ibid., 1596). Moreover they are also in a position to state whether they feel that their performance on the test constitutes a highly unrepresentative sample of their “actual” ability or not.

Low’s views are supported by Frisbie’s (1982:136) statement that “unsolicited comments and open-ended responses of students to questionnaires can provide a unique set of information to supplement other qualitative data which are used to judge the adequacy of a placement test”. Wall et al. (1994) go further by stating that the question “Do the students who take the test feel that their language has been accurately measured?” should always be asked in a study intending to measure the validity of an English language proficiency placement test.

Fulcher (1997) looks at the problem from another perspective. He states that: “If the test is not perceived to be fair by test-takers and score users, the role of the placement test within the institution is compromised” (ibid., 118). A qualitative study of this nature therefore relates not to the technical qualities of a testing instrument but to the aspects of the social consequences of the testing for the institution (Messick, 1989).

Nevertheless, although institutions and teachers are aware of the fact that their students have very clearly defined opinions about the tests that they are given they often disregard these opinions. According to Mendelsohn (1989:96), the main reason behind this policy is the fact that “we [institutions and teachers] don’t really know

how to change and improve the things that they are critical of.” A number of recent studies that have examined students’ reactions to various language proficiency tests (Low, 1982; Shohamy, 1985; Scott, 1986; Wall et al., 1994; Fulcher, 1997) provided empirical evidence for the theoretical statements presented above.

Low (1982) carried out a preliminary study on 365 first-year Arts students at the University of Hong Kong. He found that the majority of the testees were really able to state opinions in relation to test validity that were parallel to ones reached through statistical analysis. In addition, the subjects clearly stated their opinions on the content bias of the test, the tasks they were asked to perform and their own performance.

The study at Lancaster University (Wall et al., 1994) conducted on 57 foreign students revealed that subjects properly evaluated the problematic parts of the English placement test employed there. Subjects were asked to evaluate an English language placement test for academic purposes consisting of reading, writing and listening parts. Informants graded reading and writing parts as satisfactory while they said that they had doubts about the listening part. Re-examination of the test by the research team (Wall et al., 1994) illustrated that there was a real mismatch between the type of the listening sub-test and the aim of the placement test. They concluded that the instructions in the listening part should be changed.

A second pair of studies (Shohamy, 1985; Fulcher, 1997) discussed why students thought that the English language proficiency test could not measure their level of language proficiency.

In Shohamy's research (1985) carried in Israel with learners of English as a second language, 90% of the students stated that they did not feel that the language test administered in their university reflected "their language". When students were asked to explain "Why?", the majority of the informants stated that they did not see a connection between the things done on the test and their real language knowledge.

Fulcher (1997) also asked foreign students at the University of Surrey whether or not the proficiency exam employed by the institution properly measured their level of English language proficiency. Most of the students did perceive the test to be fair. On the other hand, those informants who found the test unfair mostly complained about the administration of the test. They stated that either the time was not enough or the testing environment was poor. These were the factors which effected the test scores negatively (Asher, 1990). These findings led Fulcher (1997:135) to conclude that "Within any large institution there will be logistic and administrative constraints. What is not often recognized, however, is that these constraints lead to limitations on testing, which have direct impact on the reliability of score interpretation."

CHAPTER 2

BOĞAZİÇİ UNIVERSITY ENGLISH PROFICIENCY TEST (BUEPT)

In this chapter, first, the format and contents of the English proficiency tests (i.e., previous and current), used by the Boğaziçi University English Preparatory Division will be presented and discussed. Secondly, the results of a number of studies that have been conducted to evaluate BUEPT and education in YADYOK will be presented.

2.1 Boğaziçi University English Proficiency Test (BUEPT)

BUEPT is a test constructed to be used only by BU and it is based on a needs analysis of the BU university students.

All BU entrants take the BUEPT before being allowed to start their undergraduate studies. If students are successful in the BUEPT they are permitted to continue their studies in the University. If they fail they are required to spend at least one semester but normally one year in the English Preparatory Division (YADYOK) of the University, where they receive a full-time English language course. At the end of the first semester, YADYOK students take parallel forms of the BUEPT again. Those prep students who obtain a grade below 60 (out of 100) attend English courses for another semester or a year. YADYOK students who are unable to pass the proficiency exam at the end of the academic year are not allowed to register the university and their rights to follow the preparatory classes are also sustained.

However, they can take the exams offered by BU for another academic year. Those students who are still unsuccessful at the end of the second year, are expelled from BU.

2.1.1 Background

The Boğaziçi University English Proficiency Test (BUEPT) was developed by Arthur Hughes and the English Preparatory Division (YADYOK) of Boğaziçi University (BU) within the 1982-83 academic year (see Appendix A). Hughes (1988:134) defines the BUEPT as “an English language proficiency test for academic purposes at the university level.” He also points out that the aim of the test is to “distinguish those students whose English is adequate for study at the University through the medium of English from those whose English is not” (Hughes, 1984:137).

Before Hughes developed BUEPT, that is, until the beginning of the academic year 1983-84, all students registered in YADYOK were allowed to go on to their university studies if the grade derived from their achievement tests (i.e., progress tests given throughout the year), teacher ratings and the “final proficiency test” was over 60. The progress tests and the teacher ratings contributed 70% and the final test contributed only 30% to the final passing grade.

Achievement tests were mostly prepared by prep class teachers and they differed from class to class (Hughes, 1988). On the other hand, the “final proficiency test”,

constructed by YADYOK teachers, included solely multiple choice questions which were “apparently modeled on those in the Michigan Test of English Language Proficiency and TOEFL” (Hughes, 1988:135). Besides that, some of the students who were failing on the final proficiency tests were ‘jumping’ to undergraduate studies if their progress tests grades were high. Actually, under the testing system described above, 99%* of YADYOK students were permitted to continue their undergraduate studies (Hughes, 1988:136).

However, when the students who completed their time in YADYOK were taking the Michigan Test of English Language Proficiency, as one of the requirements for completing the preparatory course, more than half of them could not reach even the “minimum score for their subject area, as indicated in the test manual” (Hughes, 1988:136). That is, if they had been in another English medium university they would have been allowed to take at most one-third of the term courses, while in BU they were taking a full load. Freshman year instructors were also complaining that incoming students’ English was not good enough to follow the lectures. That is why “a great deal of the teaching in this English medium university was in fact conducted in Turkish” (Hughes, 1988:135). Instructors were either lecturing only in Turkish or after giving their lectures in English they were breaking off from time to time to provide summaries in Turkish (Hughes, 1988).

The situation described above became worse at the beginning of 1980s, and the Senate of BU was forced to choose between either becoming a Turkish medium institution or changing the English proficiency testing system. The Turkish

* During the last two years (1998 and 1999) only 53% of the students who entered BUEPT passed it (YADYOK Testing Office, 1999).

Government and the senior members of the University decided that BU should remain an English medium university. As a first step in attempting to ensure that this happened, the Senate of the University decided that entry to undergraduate studies should depend entirely on a student's proficiency test performance. In October 1982, Arthur Hughes, a Key English Language Teaching (KELT) officer on a two-year British Council contract funded by the Overseas Development Administration came and started the development of a new English language proficiency test.

According to Hughes (1988:138) freshman year was "the critical year: if students could cope with this first year in English, they could presumably cope with the following three". That is why the new English proficiency test project started by establishing the English language needs of freshman year students. Their written work was collected and scrutinized. Academic staff teaching first year undergraduates was interviewed and administered a questionnaire. Two objectives were behind these inquiries: to discover which language skills were important for freshman year students and to detect the skills in which majority of the students had serious problems.

Results of the questionnaire revealed that listening was considered to be the most needed skill for freshman year students. It was followed respectively by reading, writing and speaking skills. Instructors indicated that there should be a level of English proficiency below which none of the students should be allowed to go on to undergraduate studies. Moreover, it was pointed out that the content of the test should be based on the 'specific language requirements' (Valette, 1977) of freshman year students at BU.

After collecting all these data, Hughes and the YADYOK team tried to prepare a test that would:

- 1) distinguish those students whose English was adequate for study at the university through the medium of English from those whose English was not (Hughes, 1984:137).

That is, the new test would be criterion-referenced. Moreover, the test would have a high enough cut off point so that those students who passed the exam would be able to follow the courses taught in English.

- 2) encourage and support the kind of teaching, particularly within YADYOK, best suited to the development of the English skills necessary for study at BU. In other words, the test should actually require the students to perform just the kind of tasks that they would meet in their first year as undergraduates (Hughes, 1988:137).

- 3) have high face validity (Hughes, 1988:137). Stated differently, the test should be seen by students and teachers as appropriate and fair.

From the academic year 1983-84 on, the new proficiency test was the only determiner of whether a YADYOK student could remain at the University or not.

2.1.2 The New Boğaziçi University English Proficiency Test (BUEPT)

As indicated in the BUEPT handbook, the test consists of three main sections - listening, reading and writing - which are further divided into two subsections within themselves (see Appendix A). The exam lasts for three and a half hours. Candidates have an hour for listening, an hour for reading and 80 minutes for writing sections. Those who get A (over 70), B (from 65 to 69), or C (from 60 to 64) pass the exam and all others who get lower marks fail (Doltaş and Sevgen, 1995). The weights of the different parts of the examination are as follows: listening 30%, reading 40%, and writing 30%. Six different scorers evaluate the listening and the reading parts of the test while four different scorers evaluate the writing part. None of the evaluators can see the marks given to the particular paper by his/her colleagues.

The listening component consists of two listening comprehension sections: *while listening* and *note taking*.

In the *while listening* part the candidates, in advance, see and have the opportunity to study the questions (from 12 to 17) for five minutes (see Appendix B). Then they listen to a talk, which is usually ten minutes long, and answer the questions while listening to it. The answers required in this part are brief; often one to two- to five-word statements.

In the *note taking* part, testees hear another ten-minute lecture and take notes while listening to it (see Appendix C). Students do not see the questions until the talk is ended. They are warned by administrators to try to catch and note only the main

points and the important examples instead of attempting to write down everything they hear (Hughes, 1984). After taking notes, the candidates are faced with questions. Students have 15 minutes to reply the questions. It is not necessary to answer in complete sentences, or write the exact words used in the lecture.

Hughes (1984) states that topics for the lectures used in the listening section are drawn from disciplines such as business administration, psychology and sociology. These areas are chosen chiefly because they are the common courses taken by most of the freshman year students at the university.

Due to the inadequacies of the available equipment and the acoustics problems of rooms where the exam is administered the lectures are delivered 'live' instead of using recordings. Hughes (1984) writes that in this way, the test becomes more authentic and closer to its purpose because nearly always in universities lectures are delivered 'live'. He claims that it is hardly possible that university instructors use recordings to present the material.

The reading component also has two parts: *scanning* and *detailed reading*. In order for a student to be considered successful on this part he/she has to do two-thirds of the questions correctly.

In *the scanning part* students are expected to scan a text up to 3,000 words as quickly as possible, searching for main or easily noticeable information. The questions (about 15 in number) in this section are typically related to important ideas

and basic facts in the passage. Students do not need to comprehend the whole text. They do not have to worry about parts of the text they have difficulty with. The important skill for this section is to find the data that the questions demand as quickly as possible (see Appendix D).

In the second, *detailed reading* section, students read a passage that is about 1500 words long (see Appendix E). Here they have to answer up to fifteen questions which are assumed to require a deeper comprehension of the passage. Questions related to the text are grouped under three headings:

1. *Comprehension questions*: Candidates are supposed to read carefully written English for understanding of main ideas and important details of an argument in a text (see Appendix E, Comprehension Section).
2. *Reference questions*: Candidates are asked to recognize the relationships in the text (e.g., "Identify the antecedent referent of pronoun", Hughes, 1984:140). The idea is to see whether students can recognize the way in which such pronominal devices refer to information elsewhere in the text (see Appendix E, Referrals).
3. *Vocabulary*: Candidates are given four to six words and they are asked to find a single word in the passage between the pointed out lines, which has the same meaning. Candidates are warned that "the word in the passage may have different endings e.g., -ing, -ed, etc." (BUEPT Manual 1997:4). The aim is to check whether examinees can derive the meaning of unfamiliar vocabulary from the context (Hughes, 1984) (see Appendix E, Vocabulary).

The passages used in the reading component of BUEPT are taken from the university textbooks, the subject areas being similar to those chosen for the listening component texts.

The writing component: Here students have to write two one-page compositions usually about pros and cons of a particular topic (see Appendix F, The Argumentative Essay) or they have to compare and contrast two things (e.g., Compare and contrast life in big cities and villages). The students are reminded that they should demonstrate that their level of English is enough for university studies and because of this, they should not, for instance, use only short and simple sentences.

Hughes (1984:140) states that some guiding points are presented to students since more than assessing creativity, the intent of each task is to measure whether candidates' writing abilities are adequate for study at BU. He gives the following writing component example:

'Discuss the following proposed measures intended to increase the number of foreign tourists coming to Turkey.

(1) More/ better advertising and/ or information (where? what form should it take?)

(2) Improved facilities (hotels, transportation, communication etc.)

(3) Training of personnel (guides, hotel managers etc.)

Hughes (1984:142) presents the following scoring system for the writing component of BUEPT:

<u>NS</u>	: <i>Educated Native Speaker Standard</i>
<u>NS-</u>	: <i>Very Close to Native Speaker Standard</i>
<u>MA</u>	: <i>Clearly More Than Adequate</i>
<u>MA?</u>	: <i>Probably More Than Adequate</i>
<u>A</u>	: <i>Adequate for Study at BU</i>
<u>D</u>	: <i>Doubts about Adequacy</i>
<u>NA</u>	: <i>Not Adequate</i>
<u>FBA</u>	: <i>Far below Adequacy</i>

Those students whose compositions fall in the first five groups get passing marks on the writing part of the exam while the compositions in following three groups fail (see Appendix G for samples of essays graded differently, Hughes, 1984:145-147). Hughes (1984) pointed out that two sources were used to establish the standards for composition evaluation:

- (1) The written work of undergraduate students: Some university instructors supplied a sample of learners' papers with their notices about the adequacy of the students' English for university study.
- (2) Ideas, thoughts, and experiences of the teachers in YADYOK, who themselves had graduated from BU.

2.1.3 Comments on the Components of BUEPT

Hughes's (1984, 1988) description of the components of BUEPT may raise a number of concerns with regard to the validity and reliability of the test.

First of all, the fact that the texts used in both the reading and the listening sections of BUEPT are drawn from only three disciplines (i.e., business administration, psychology and sociology) may raise the problem of content validity. Restricting the test topics to those three areas may lead to the following questions: "To what extent is BUEPT really able to reflect the representative language that all examinees might encounter in the criterion (i.e., BU) academic setting?", "Does BUEPT overlook the performance of engineering, science and philology students?" and "Is BUEPT biased towards the students in Business Administration, Psychology and Sociology Departments?"

In order to answer the questions posed above empirical studies comparing the academic performance of the students in Business Administration, Psychology and Sociology Departments with the academic attainment of the students in other departments need to be carried out.

A second issue concerns the construct validity of the listening, reading and writing components of BUEPT. Bachman (1990) points out that before trying to measure something it has to be properly and fully identified and defined. Harmer (1993:183) in his discussion of reading and listening skills states that these skills have at least six basic dimensions, i.e., predictive skills, skills for extracting specific and detailed

information, skills for getting the general picture, skills needed for recognizing function and discourse patterns, and skills for deducing meaning from the context. All these have to be possessed by readers/listeners in order for them to be able to understand the content of what they see/hear. However, in BUEPT's case the constructs of reading and listening for academic purposes are limited to only two dimensions – scanning and detailed reading, and while listening and note taking respectively. Brown (1996:251) points out that in cases such these, the mismatch between the test scores and the criterion behavior comes not from measuring different abilities, but from measuring only limited aspects of the criterion ability.

Hughes (1984, 1988) consistently states that BUEPT was prepared to measure whether or not students possess the skills required to cope in an academic setting. Nevertheless, in the writing part (similar to the skills requirements in the reading and listening parts) examinees are always asked to write the same types of essays – the argumentative and the comparison-contrast essays. Since it is commonly accepted that academic writing is not limited to only these two types of writings, there may be a concern about the construct validity of the writing section of BUEPT as well.

In addition, nothing is said about the importance of style, structure and/or the usage of graphic symbols (e.g., punctuation symbols, capitalization, indentation, italics, underlining) in the marking scheme presented by Hughes (1984:42, see p. 48). As it is well known, academic writing has its strict rules that are expected to be followed by academicians and students.

A third issue concerns the listening component of BUEPT. The fact that the “talks” in listening sections are delivered ‘live’ may have implications for the reliability of the test. As Cronbach (1971:449) states “every aspect of the setting in which the test is given and every detail of the procedure may have an influence on performance and hence on what is measured.” The ‘thorny problem of the accent’ (Gue & Holdaway, 1973) and the speed of reading of the lecturer (Wall et al., 1994), for instance, may be factors affecting both the success of examinees and the reliability of the exam.

The final concern with regard to BUEPT is the present populations of the test-takers. As explained previously, the content and the skills that are assessed on the exam are arranged according to the requirements of the freshman year courses taught at BU. However, BUEPT is also given to “would be” BU graduate students and outsiders (e.g., members of the staff of different banks). An obvious question that has to be answered then is: “Can BUEPT measure the level of English proficiency of those examinees since this was not the original intention?”

Anastasi and Urbina (1997:138) point out that it is essential to specify the features of the group for whom a test is prepared. They claim that the same test may measure different functions when given to individuals who differ in age, sex, educational level, and any other relevant characteristic. Persons with different experiential backgrounds, for example, may utilize different work methods to solve the same test problem. Consequently, a test could have validity in predicting a particular criterion in one population and little or no validity in another. That is why, a test designed for use with a particular population should cite appropriate data on population generalizability in its technical manuals.

2.2 Review of BUEPT and YADYOK Studies

To the knowledge of the researcher, there are only three studies that have been carried out to evaluate the BUEPT or English education in YADYOK.

The first known study was conducted by Hughes in 1983. The BUEPT scores of 190 advanced BU students were correlated with their scores on the Michigan Test of English Language Proficiency (MTELP). The calculations ended up with a strong relationship between the two sets of data (r varied from .70 to .84). Depending on this high correlation, Hughes (1988) concluded that the two tests measured the same construct - the level of English language proficiency of university students, though in different ways (i.e., MTELP is a multiple choice or "objective" test while BUEPT is a "subjective" exam as defined by Pilliner, 1968).

The second survey was carried out by Doltaş and Sevgen (1995). The aim of the study was to examine whether or not there was a correlation between BUEPT and I-TOEFL scores. The participants in the study were 64 freshman year undergraduate and 66 graduate and transfer BU students, who took both tests within the same semester. The analysis of undergraduate students' data revealed that in contrast to the researchers' expectations, a high score on I-TOEFL (e.g., 600-677) did not guarantee an "A" on BUEPT. Among the six students who had both high I-TOEFL scores (between 600-660) and passed the BUEPT at first try, only two got "A" on the latter test. On the other hand, all of the ten subjects who attended prep classes for one semester, and had TOEFL scores ranging between 500 and 573, also passed the

BUEPT with “A”. These findings led Doltaş and Sevgen (1995) to argue that the determining factor which indicated the students’ success on BUEPT was not their level of English proficiency but their knowledge of test techniques.

When the graduate and transfer students’ data were examined it was detected that among the 66 candidates who had an I-TOEFL score of 550 or above, only one passed the BUEPT with “B”. The rest of the students, except for two who failed, passed the exam with “C”. The TOEFL scores of those who got the same grade on BUEPT (i.e., “C”) ranged from 550 to 677. Considering these data, researchers concluded that A, B, C ranges of BUEPT were not sensitive to the different levels of English language proficiency of subjects in the study.

The BU Alumni Association (BUMED) conducted a widescope study trying to measure the “performance of BU”. One of the aims of this survey was to discover how prep class students evaluate the education given in YADYOK. When asked whether or not their expectations related to the teaching in YADYOK were fulfilled, 83,2% of prep students said that their expectations were either only partially fulfilled or not fulfilled at all. When asked “Why?” more than half of those informants (55.4%) stated that they were dissatisfied by the way English is taught in YADYOK. Another group of subjects (8%) noted that they were disappointed by the low level of English proficiency, both of YADYOK teachers (2.9%) and incoming students (5.4%).

Considering the conflicts between the BUEPT designer's aims and the results of the last two studies discussed above, it is believed that the BUEPT may need closer examination. The way the test is employed (i.e., a screening tool for detecting students with deficient English language skills, which might block students' success in undergraduate studies) and the population affected the most by it (i.e., the students) makes this inquiry of the predictive and face validity of BUEPT reasonable.

CHAPTER 3

METHODOLOGY

3.1 Purpose Of The Study

The purpose of the current study was to answer the following research questions:

1. Is there a relationship between the time students spend in YADYOK and their success on BUEPT?
2. Is the Boğaziçi University English Proficiency Test (BUEPT) a valid predictor of the first semester, second semester and overall freshman year academic success of Boğaziçi University (BU) students in Foreign Language Education (FLED), Western Languages and Literatures (LL) and Translation and Interpretation (TRANS) Departments? Why or why not?
3. How do the students in FLED, LL, and TRANS Departments evaluate the BUEPT as a valid reflector of their level of academic English proficiency and as a predictor of their future academic success?

It was expected that in answer to the first research question, a positive relationship would be found between the time students spent in YADYOK and their success on

BUEPT, since courses in YADYOK are designed to prepare students specifically for the BUEPT (Hughes, 1988; Onat, 1997).

As explained previously, BUEPT is intended to determine whether the level of English proficiency of incoming BU students is sufficient to follow university courses or not (Hughes, 1988). Furthermore, Weir (1988:45) claims that a student's proficiency in English determines to "what degree he or she will benefit from, and contribute to, the course of study". Moreover, the available empirical evidence cited earlier suggested that academic achievement in social sciences is affected more by the level of English language proficiency, than is the academic achievement in other fields of study (Light et al., 1987).

Considering the points mentioned above and the facts that the present study was carried out in an English-medium university and all the subjects were from departments closely related with the English language, it was expected that in answer to the second research question a positive relationship would be found between the BUEPT results and the first semester, second semester and overall freshman year academic success of the subjects in the research.

The GPA was chosen as a criterion for academic achievement due to two reasons. First, it was in line with most of the studies reviewed in Chapter 1. Second, the GPA is the most important factor determining the future academic life of students at BU. According to the 1998 BU Student Booklet: "All students in any undergraduate program at BU should have a GPA which is at least 2.00. If within two successive semesters a student's GPA is below 2.00, he/she is considered to be a "repeating"

student. "Repeating" students cannot take new courses in the following terms. Beginning with courses in which they failed (received F), they have to repeat courses which they passed with DD and DC until their GPAs become at least 2.00" (Boğaziçi Üniversitesi El Kitabı, 1998:31).

It was decided to correlate BUEPT results particularly with the first semester, second semester and overall freshman year GPAs of the students, because the content of the test was based on the "specific language requirements" of the freshman year students (Hughes, 1988). Furthermore, as Hughes (1988) and Enginarlar (1984:119) pointed out it is the most important year at the university education, failure at the end of which "may very well be, among other things, causally related to language proficiency".

Although it was less easy to predict the answer to the third research question, it was expected that the exam would be "seen by students as appropriate and fair" (Hughes, 1988:137). It was hoped that the results obtained from this part of the study would lead to actionable administrative and pedagogical recommendations. That is why, the data collection procedures were designed such that detailed information could be obtained about the attitudes and thoughts of subjects towards the BUEPT.

3.2 Subjects

A total number of 422 (332 female and 89 male) BU students from Foreign Language Education (FLED) (207), Western Languages and Literatures (LL) (114)

and Translation and Interpretation (TRANS) (101) Departments, who were admitted to the university between 1992 and 1997, participated in this study. The students' ages ranged from 20 to 35.

A background survey indicated that the target population was rather homogeneous. Nearly all students (409 out of 422) had had instruction in English in junior and senior high school for five to seven years (see Appendix H). They were either private high school (e.g., Robert College and Üsküdar American College) or Anatolian public high school (e.g., Edirne Anadolu Öğretmen Lisesi and Balıkesir Anadolu Lisesi) graduates, where the medium of instruction is English. The English curriculum for these institutions is designed by the Ministry of Education to bring students who graduate from high school, to an advanced level of English proficiency. The other thirteen students were public high school graduates but they explained that they had taken private lessons and attended private university preparatory courses, which helped them to improve their level of English proficiency.

As it is known, there is a central matriculation system in Turkey. That is, all university candidates have to take the University Entrance Exam prepared and administered by the Student Selection and Placement Center (ÖSYM) set in Ankara. In order to be able to enter FLED, LL and TRANS Departments of BU, candidates have to pass the English Test of the University Placement Test (ÖYS). The ÖYS English Test consists of 75 multiple choice questions, usually based on the topics included in the curriculum of Anatolian and Private High Schools. The fact that all subjects took and did at least 70 out of 75 questions on the English Test of the ÖYS

(between 1992 and 1997) was accepted as further evidence that all subjects were more or less at the same level of language proficiency.

Thus, it can be said that factors that might have affected the results of the current study such as the subjects' pre-entry level and previous educational background were controlled in this study as much as possible.

3.3 Data Collection

The following data collection procedures were used in the study: A) Obtaining statistical data from the BU Registrar's Office; B) Student Questionnaire.

A) Statistical Data

The statistical data obtained from the BU Registrar's Office included the BUEPT passing grades and the GPAs of 422 students in FLED, LL and TRANS Departments. The available data also allowed the identification of: (1) The academic major of the subjects; 2) The number of subjects who passed the BUEPT at first try; 3) How many semesters each of the subjects spent in preparatory classes; 4) The subjects' GPAs throughout the different semesters in the freshman year.

B) Questionnaire

A questionnaire used by Wall, Clapham and Alderson (1994) in their study for evaluating the English Proficiency Test at Lancaster University formed the basis of much of the questionnaire employed in the current research (see Appendix H). However, some items were rewritten, and others were replaced by new ones in order to make the questionnaire more suitable for its particular context.

The questionnaire was chosen as a data collection instrument in this study because, as Harris (1983) stated, the only way to find out about face validity that is concerned with what students think of the test is by means of a questionnaire. In addition, Low (1985) expressed that the best way to collect a large amount of data quickly was a standardized questionnaire. Furthermore, as Nunan (1992:143) pointed out, data from a standardized questionnaire is more “amenable to quantification”. According to him, these data can be readily quantified and analyzed, particularly if there is access to computer statistical packages. The types of questions were designed in accordance with Low’s suggestions (Low, 1985:165). He indicates that if the data are going to be collected from a large population, the questionnaires should mostly comprise close-ended responses followed by places for open-ended responses in order to elicit data with respect to the actual ideas and feelings of the subjects.

The questionnaire consisted of two parts (see Appendix H). The first part of the questionnaire included questions related to the background of the subjects: age, gender, educational background, success on the ÖYS.

The questions in the second part of the questionnaire were itemized according to the second foci of the study: overall perception and evaluation of the BUEPT by FLED, LL and TRANS students.

While trying to make the questionnaire as detailed as possible, there was a need to limit the number of items to those which could be tackled by students within a maximum of 20 minutes. This limit was determined during a personal conversation with the target instructors, who told the researcher that they might allow 20 minutes of their lessons to be taken up by the researcher.

3.4 Piloting

A pilot study of the questionnaire (adapted from Wall et al., 1994) was conducted with 40 FLED students. Besides trying to detect the problems related to the format and content of the questionnaire, it was also aimed at determining how much time was needed to finish it. Subjects in the pilot group were asked to write their comments and suggestions related to each of the items while answering the questions.

Students were able to complete the questionnaire within 25 minutes. This was considered as a good indicator of the time within which the questionnaire could be completed. Moreover, no significant problems were identified from the piloting procedure. All questions were answered fully by all students, and except for two, none of them wrote negative comments related to the items. This was accepted as

evidence that the questionnaire was not too long and that the questions were not boring or irrelevant to the interests of the target population.

Later, the results of the pilot study and the finalized questionnaire form were examined by three instructors in the FLED Department. Since no significant problems were identified by them, the questionnaire was administered to the target population.

3.5 Data Analysis

Two different types of data were obtained as a result of the data collection procedures: A) Quantitative data of the Registrar's Office and of the close-ended questions of the questionnaires; B) Qualitative data of the open-ended questions of the questionnaires.

A) Statistical Data

First of all, in order to make the calculations and comparisons throughout the study more meaningful and reliable the BUEPT passing grades were reconciled with the academic grading system at BU (see Table 3.1). Reconciliation was based on the information presented at the 1998 BU Student's Booklet (1998:27).

Table 3.1
Reconciled BUEPT scores

BUEPT passing grade	Expected GPA
C	2.00-2.49
B	2.50-2.99
A	3.00-4.00

The Statistical Package for Social Sciences (SPSS) was used for the analysis of the reconciled quantitative data. In accordance with the objectives of the study, descriptive statistics, analysis of variance and Spearman Rank Order Correlation Coefficients between the BUEPT scores and the first semester, second semester, and overall freshman year GPAs of the subjects were calculated.

B) Questionnaire

In order to synthesise the qualitative data obtained from the questionnaire, a “key word analysis” (Nunan, 1992) was conducted. According to Nunan, with this procedure, the statements made by subjects can be generated into categories without distorting or misinterpreting the available information. As a second step, frequency count and percentage computation of all of the responses on the questionnaire were performed.

CHAPTER 4

RESULTS AND INTERPRETATIONS

The findings of the study with their interpretations are presented in this chapter. In Part A, the findings related to the first and second research questions “Is there a relationship between the time students spend in YADYOK and their success on BUEPT?” and “Is the Boğaziçi University English Proficiency Test (BUEPT) a valid predictor of the first semester, second semester and overall freshman year academic success of Boğaziçi University (BU) students in Foreign Language Teaching (FLED), Western Languages and Literatures (LL), and Translation and Interpretation (TRANS) Departments? Why or why not?” will be discussed. Then, in Part B, the findings related to the study’s second question “How do the students in FLED, LL and TRANS Departments, evaluate BUEPT as a valid reflector of their level of English proficiency and as a predictor of their future academic success?” will be presented. What the subjects of the study think of the test and the preparatory classes will also be explained.

PART A: The Answer To Research Questions 1 and 2

The available data were analysed under four sets: (1) G1 – the group of students who passed BUEPT at first try and ended the freshman courses without any prep English instruction; (2) G2 – the group of students who passed BUEPT after a semester of preparatory school; (3) G3 – the group of students who passed BUEPT after two

semesters of preparatory school; (4) G – the total group of students (including all subjects in the study).

To obtain a comprehensive and yet structured overview of the relationship between BUEPT scores and academic success, the analysis was done in three steps. Firstly, the distribution of BUEPT scores according to the time students spent in the English Preparatory Division (YADYOK) of Boğaziçi University was scrutinized. Secondly, the associations between the BUEPT scores and the first semester, second semester and overall freshman year GPAs of the students were examined. Finally, an in-group statistical analysis intending to answer the “Why or why not?” part of the second research question was conducted.

4.1 Distribution Of BUEPT Scores According to the Time Students Spent in the English Preparatory Division (YADYOK) of Boğaziçi University

The initial part of the analysis in this section was started by scrutinizing the distribution of BUEPT scores according to the time students spent in YADYOK. It was aimed at answering the following question: “Is there a relationship between the time spent in YADYOK and the success on BUEPT?”

When the distribution of BUEPT results for the three groups was closely examined, it was seen that the answer to this question was “Yes”. A positive relationship was found between the time spent in YADYOK and examination grades.

As is displayed in Figure 4.1 and Table 4.1, more than half of the students (55%) in G1 passed BUEPT with the lowest possible grade “C”, one-fourth passed it with “B”, and only 20% were able to pass the exam with “A”. In G2, after only one semester of prep classes, the number of those who passed the exam with “A” and “B” increased and became respectively 22% and 36%, while the number of those who passed with “C” decreased 13% and became 42% (from 55% to 42%). In G3, there were more students who passed the exam with “A” (34%) and “B” (34%) than those who passed it with “C” (32%).

Figure 4.1
Distribution of BUEPT scores according to the time students spent in YADYOK

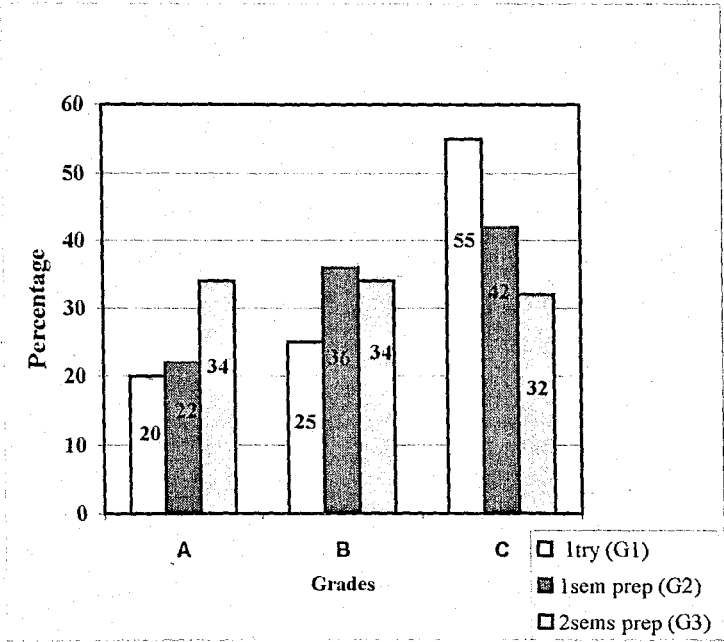


Table 4.1
Distribution of BUEPT scores according to the time students spent in YADYOK

	A		B		C		Total	
	Number	%	Number	%	Number	%	Number	%
First Try (G1)	36	20	44	25	97	55	177	42
One Sem. Prep (G2)	20	22	32	36	38	42	90	21
Two Sems. Prep (G3)	52	34	52	34	51	32	155	37
Total (G)	108	26	128	30	186	44	422	100

Later, a chi-square analysis was conducted to examine whether the difference in pass rates among the three groups was statistically significant. The outcomes of the analysis showed that the students who received instruction in YADYOK performed significantly better on the exam than those who did not receive such instruction ($\chi^2 = 17.72$; d.f. = 4; $p < .01$). G3, the group that spent a whole year in YADYOK performed significantly better than the other two groups (G1 and G2), and G2 performed significantly better than G1. Thus, it can be said that YADYOK, where the courses are designed to “prepare students specifically for this [the BUEPT] exam” (Onat, 1997:23), fulfills its requirements and prepares the students for the BUEPT.

4.2 Statistical Analysis of the Relationship between the BUEPT Scores and the First Semester, Second Semester and Overall Freshman Year GPAs

In this section the relations between BUEPT passing grades and the first semester, second semester and overall freshman year GPAs of the students are presented.

4.2.1 Descriptive Statistics/ Comparison of Means

Considering the first finding of the study and the statement that BUEPT is a test requiring students to “perform just the kinds of tasks that they would meet in their first year as undergraduates” (Hughes, 1988:143), it was hypothesized that there would be a positive relation between BUEPT results and the first semester, second semester and overall freshman year GPAs of the subjects.

Briefly, the Null Hypothesis tested was: “There is a significant correlation between the BUEPT results and the first semester, second semester and overall freshman year GPAs of English major students (i.e., students in FLED, LL and TRANS Departments) at Boğaziçi University”.

To test the Null Hypothesis, descriptive statistics and One-way analysis of variance were used.

Table 4.2
Descriptive Statistics

	First Try (G1)			One Sem. Prep (G2)			Two Sems. Prep (G3)			Total (G)		
	1 st Sem	2 nd Sem	Year	1 st Sem	2 nd Sem	Year	1 st Sem	2 nd Sem	Year	1 st Sem	2 nd Sem	Year
Mean (M)	2.39	2.4	2.4	1.93	2.19	2.06	2.04	2.20	2.12	2.16	2.28	2.22
Mode	2.5	2	2.25	1.5	1	1.4	2.5	2	2	2.5	2	2.25
Median	2.5	2.46	2.44	1.93	2.19	2.1	2.08	2	2.13	2.25	2.29	2.25
Range	3.9	3.8	3.34	3.42	3.36	3.26	3.5	3.6	3.33	4	3.8	3.48
SD	.651	.638	.535	.676	.689	.586	.757	.640	.580	.724	.657	.580

Descriptive statistics, which were expected to give the numerical representation of the academic performance of students during their first year as undergraduates, were calculated. As can be seen in Table 4.2, G1, the group with the lowest achievement level on BUEPT, was the most successful group throughout the freshman year ($M_{1SEM1} = 2.39$; $M_{1SEM2} = 2.40$; $M_{1YEAR} = 2.40$). G3 was the group that followed G1 in terms of academic achievement ($M_{3SEM1} = 2.04$; $M_{3SEM2} = 2.20$; $M_{3YEAR} = 2.12$), and G2 was the group with the lowest first semester, second semester and overall freshman year mean ($M_{2SEM1} = 1.93$; $M_{2SEM2} = 2.19$; $M_{2YEAR} = 2.06$).

Besides this general picture, Table 4.2 presents important details related to the academic performance of the three groups throughout the different semesters of the

freshman year. While first and second semester means of G1 were almost identical ($M1_{SEM1} = 2.39$; $M1_{SEM2} = 2.40$), the second semester means of both G2 and G3 were higher than their first semester means ($M2_{SEM1} = 1.93$ vs. $M2_{SEM2} = 2.19$; $M3_{SEM1} = 2.04$ vs. $M3_{SEM2} = 2.20$). Moreover, the second semester mean of G2 was significantly higher than its first semester mean ($p < .05$).

A possible explanation for the improvement of the grades of G2 might be the fact that they took the freshman year courses in the “wrong” order. That is, they took second semester courses (and really had a hard time coping with them), before taking the first semester courses.

Table 4. 3
ANOVA – Scheffe – Comparison of Means

	Compared Sets		Mean Difference	Std. Error	Sig.
FIRST SEMESTER	G1	G2	.4635**	.009	.000
		G3	.3562**	.077	.000
	G2	G3	.1073	.092	.510
SECOND SEMESTER	G1	G2	.1928*	.072	.028
		G3	.1927*	.072	.028
	G2	G3	.0012	.086	.998
END OF THE FRESHMAN YEAR	G1	G2	.3280**	.073	.000
		G3	.2747**	.062	.000
	G2	G3	.0611	.075	.775

*The mean difference is significant at the .05 level.

**The mean difference is significant at the .000 level

One-way analysis of variance was used to check if there were any significant differences among the means of the three groups. As displayed in Table 4.3, G1 had a significantly higher mean than both G2 and G3 at the end of the first semester,

second semester and freshman year. However, none of the differences between the means of G2 and G3 were found to be significant. That is, G1, whose achievement level was significantly lower than the other two groups (G2 and G3) on BUEPT, performed significantly better than those two groups during the whole freshman year.

According to these findings, the hypothesis that was put forward at the beginning of this section was rejected. The results imply that on group bases, the success on BUEPT tended to accord with lower academic achievement during the freshman year. Moreover, attending preparatory classes resulted in better performance on BUEPT, but did not lead to better academic attainment. Spending one more semester in YADYOK did not help the subjects in G3 to achieve GPAs that were statistically higher than the first semester, second semester and overall freshman year GPAs of subjects in G2, either.

Descriptive statistics and the analysis of variance allowed the rejection of the Null Hypothesis but these results were only informative in terms of giving the strength and direction of the association between the BUEPT results and the freshman year GPAs. To get more insight into the degree and significance of the relationship between the criterion measures (i.e., BUEPT scores and first semester, second semester and overall freshman year GPAs) the Spearman rank order correlation coefficients were calculated.

4.2.2 Spearman Rank Order Correlation Coefficients

In this section the calculated Spearman rank order correlation coefficients are presented and discussed.

Table 4. 4
Spearman Rank Order Correlation Coefficient

	First Try (G1)			One Sem. Prep (G2)			Two Sems. Prep (G3)			Total (G)		
	GPA1	GPA2	Year	GPA1	GPA2	Year	GPA1	GPA2	Year	GPA1	GPA2	Year
BUEPT Scores	-.232*	-.110	-.199*	-.391**	-.338*	-.396**	-.401**	-.310**	-.452**	-.248**	-.161**	-.224**
Sig.	.002	.144	.008	.001	.000	.000	.000	.000	.000	.000	.000	.000
N	177	177	177	90	90	90	155	155	155	422	422	422

*Correlation is significant at the .01 level (2-tailed)

**Correlation is significant at the .000 level (2-tailed)

As indicated in Table 4.4, all of the Spearman rank order correlation coefficients were negative, ranging from -.110 to -.452, with eight being significant at the .000 level, and only one not reaching significance at the .01 level.

Analysis also revealed that BUEPT predicted second semester GPAs of each of the three groups better than their first semester and overall freshman year GPAs. Moreover, it was found that as the period of time spent in YADYOK increased, the predictive feature of BUEPT decreased. Among the three groups, the freshman year academic success of G1 (i.e., the group of students who passed the BUEPT at first try) was predicted the best ($r_{1GPA1} = -.232$; $r_{1GPA2} = -.110$; $r_{1YEAR} = -.199$), while the first year performance of G3 (i.e., the group of students who spent a whole year in preparatory classes) was predicted the least successfully ($r_{3GPA1} = -.401$; $r_{3GPA2} = -.310$; $r_{3YEAR} = -.452$). The correlation between the BUEPT scores and the first

semester, second semester and overall freshman year GPAs of G2 were as follows:

$r_{2GPA1} = -.396$; $r_{2GPA2} = -.338$; $r_{2YEAR} = -.396$.

This analysis, backed up by the data in Table 4.2 and Table 4.3, would seem to suggest two key findings. First, BUEPT is not a valid predictor of the first semester, second semester and overall freshman year academic success of English majors at BU. Second, the longer the students attend preparatory classes, the weaker the relationship between BUEPT score and their future academic performance becomes.

4.3 Within Group Analysis

In order to find possible explanations for the negative correlation coefficients obtained between BUEPT scores and the freshman year GPAs, a within group analysis was conducted. First, three Contingency Tables showing the distribution of the first semester, second semester and overall freshman year GPAs of students according to BUEPT passing grade – A, or B, or C – were compiled. After that, an analysis of variance, comparing the GPAs of students who passed BUEPT with different grades in each of the three groups – G1, G2, G3 – was conducted.

4.3.1 Contingency Tables

Contingency Tables 4.5, 4.6 and 4.7 show the distribution of the GPAs of the students at the end of the first semester, second semester and freshman year,

respectively. To perceive the similarities and/or differences in the academic performance of students during the different semesters of the freshman year, data displayed in these tables will be analyzed and compared.

4.3.1.1 First Semester Contingency Table

Table 4. 5 Contingency Table – FIRST SEMESTER

	FIRST TRY (G1)								ONE SEMESTER PREP (G2)								TWO SEMESTER PREP (G3)							
	Cs		Bs		As		TOTAL		Cs		Bs		As		TOTAL		Cs		Bs		As		TOTAL	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
0.00-1.99	26	27	10	23	6	17	42	24	27	71	16	50	6	30	49	54	35	68	19	36	12	23	66	42
2.00-2.49	26	27	12	27	3	8	41	23	5	13	8	25	7	35	20	23	7	14	18	35	11	21	36	23
2.50-2.99	34	35	13	30	14	39	61	34	4	11	6	19	6	30	16	18	5	10	13	25	17	33	35	23
3.00-4.00	11	11	9	20	13	36	33	19	2	5	2	6	1	5	5	5	4	8	2	4	12	23	18	12
Total	97		44		36		177		38		32		20		90		51		52		52		155	
Column Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

First the “Total” columns of Table 4.5 were examined. It was noticed that more than half (54%) of the students in G2 and 42% of the students in G3 failed (i.e., had GPAs below 2.00) at the end of the first semester of the freshman year. On the other hand, 12% of G3 and only 5% of G2 students reached “honor” GPAs (i.e., GPAs above 3.00). Compared to the other two groups, G1 had higher academic achievement during the first semester of the freshman year. Even though 24% of the students in this group failed, 53% of them had GPAs above 2.50 (19% of whom were “honor” students).

When the distribution of the first semester GPAs in relation to the BUEPT passing grade was analyzed, the outcome was not surprising. It was found that those students who passed the exam with "C" had the largest percentage of failures (55%), and those who passed it with "A" had the lowest percentage of failures (23%). The percentage of failing students among those who passed with "B" was 36%.

The most unexpected fact exhibited in Table 4.5 was the really high percentage of failing "C" students in G2 (71%) and G3 (68%). That is, every third student in those two groups had a GPA below 2.00 at the end of the first semester. Among the "C" students in G1, 27% finished the first semester of the freshman year with GPAs between 0.00 and 1.99.

Another salient fact displayed in Table 4.5 is the high rate of failures among those who passed the BUEPT with "B". It was expected that those students would mostly have GPAs between 2.50 and 2.99 (see Table 3.1). Nevertheless, half of the "B" students in G2 (50%), more than one-third of those in G3 (36%) and 23% of the "B" students in G1 had GPAs below 2.00 at the end of the first semester.

Among the students who passed the exam with "A", those in G2 were the least successful ones. While only 5% of them reached GPAs above 3.00, one-third (30%) failed and 35% got GPAs between 2.00 and 2.49. When the first semester GPAs of "A" students in G3 were scrutinized, it was noticed that while one-fourth (24%) of them failed, 56% got GPAs above 2.50. At the end of the first semester of the freshman year, the most successful of all students were those who passed the BUEPT

with “A” at first try. Although 17% of those students failed, 36% of them got “honor” GPAs.

4.3.1.2 Second Semester Contingency Table

Table 4. 6. Contingency Table – SECOND SEMESTER

	FIRST TRY (G1)								ONE SEMESTER PREP (G2)								TWO SEMESTER PREP (G3)							
	Cs		Bs		As		TOTAL		Cs		Bs		As		TOTAL		Cs		Bs		As		TOTAL	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
0.00-1.99	20	21	10	23	3	8	33	19	20	53	9	28	2	10	31	35	16	31	16	31	9	17	41	26
2.00-2.49	34	35	13	30	10	28	57	32	8	21	9	28	5	25	22	24	24	47	23	44	13	25	60	39
2.50-2.99	29	30	16	36	14	39	59	33	8	21	11	35	10	50	29	32	8	16	10	19	18	35	36	23
3.00-4.00	14	14	5	11	9	25	28	16	2	5	3	9	3	15	8	9	3	6	3	6	12	23	18	12
Total	97		44		36		177		38		32		20		90		51		52		52		155	
Column Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

When the “Total” columns of the second semester Contingency Table were examined it was noticed that G2 (35%) was the group with the highest rate of failures, followed by G3 (26%), and G1 (19%). However, it should also be stated that more students in G2 (41%) than in G3 (35%) had GPAs above 2.50 at the end of the second semester of the freshman year. The percentage of students with GPAs higher than 2.50 in G1 was 49%.

Besides that, at the end of the second semester the difference in academic attainment among the three groups decreased. While at the end of the first semester 19% of the students in G1 and only 5% of the students in G2 had “honor” GPAs, at the end of

the second semester the rate of “honor” students in G1 decreased (19% at the end of the first semester, 16% at the end of the second semester), while in G2 increased (5% at the at the end of the first semester, 9% at the end of the second semester). No difference was found in the rate of “honor” students in G3 (12% at the end of both semesters).

The examination of the distribution of GPAs according to the BUEPT passing grade revealed that those who passed the exam with “C” or “B” failed much more often than those who passed the test with “A”. While 35% of all “C” and 28% of all “B” students were failing, only 9% of those who passed the BUEPT with “A” had GPAs below 2.00 at the end of the second semester of the freshman year.

Among the “C” students in the three groups – G1, G2 and G3 – those in G2 were the least successful ones. More than half of them (53%) had GPAs between 0.00 and 1.99 at the end of the second semester. One-third (31%) of the students in G3 and 21% of the students in G1 had GPAs within the same range. On the other hand, while 14% of the students in G1 got “honor” GPAs, only 5% of G2 and 6% of G3 students were able to reach GPAs above 3.00 at the end of the second semester.

Table 4.6 also presented a number of noteworthy problems related with “B” and “A” students in G3. Those students who passed the BUEPT with “B” after two semesters of prep were failing more often than both the “B” students in the other two groups (23% in G1, 28% in G2 and 31% in G3) and the “C” students in G1 (21%). Besides that, they were failing as often as the “C” students (31%) in the same group.

Similarly, G3 students who passed the BUEPT with “A” failed more often than the students who passed the exam with the same grade in G1 and G3 (8% in G1, 10% in G2 and 17% in G3) at the end of the second semester of the freshman year. So, contrary to expectations, passing the BUEPT with “B” or “A” did not guarantee a successful academic life for students in G3.

4.3.1.3 Freshman Year Contingency Table

Table 4. 7: Contingency Table – FRESHMAN YEAR

	FIRST TRY (G1)								ONE SEMESTER PREP (G2)								TWO SEMESTER PREP (G3)							
	Cs		Bs		As		TOTAL		Cs		Bs		As		TOTAL		Cs		Bs		As		TOTAL	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
0.00-1.99	27	28	10	23	3	8	40	23	23	60	11	34	4	20	38	42	31	61	21	40	9	17	61	40
2.00-2.49	29	30	18	40	10	28	57	32	9	24	12	38	9	45	30	33	13	25	24	47	7	13	44	28
2.50-2.99	31	32	10	23	12	33	53	30	4	11	8	25	5	25	17	19	6	12	7	13	28	55	41	26
3.00-4.00	10	10	6	14	11	31	27	15	2	5	1	3	2	10	5	6	1	2	0	0	8	15	9	6
Total	97		44		36		177		38		32		20		90		51		52		52		155	
Column Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

The “Total” columns in Table 4.7 show that at the end of the freshman year, 23% of the students in G1, 42% of the students in G2 and 40% of the students in G3 had GPAs below 2.00. Stated differently, a pretty high number of the subjects in the current study became “repeating” students at the end of their first year as undergraduates. Those students were not allowed to take new courses in the following semester, and they had to repeat the courses that they had taken before

until their GPAs became at least 2.00. On the other hand, a small percentage of subjects in each group reached “honor” GPAs (15% in G1, 6% in G2 and 6% in G3).

When the distribution of GPAs in relation to the BUEPT passing grade was analyzed, it was found that there was a pattern across the different semesters. That is, those students who passed the exam with “C” were failing the most (50%), and those who passed it with “A” were failing the least (15%) during the freshman year, independently of whether or not they attended prep classes. The percentage of failing students among “Bs” was 33%.

At the end of the freshman year, two-thirds of the “C” students in G2 (60%) and G3 (61%) had GPAs below 2.00. Compared with them, only 28% of those students who passed the BUEPT at first try failed. On the other hand, while 42% of the students in G1 got GPAs above 2.50, only 16% of G2 and 14% of G3 students reached the same GPAs.

Another significant problem detected by the analysis and displayed in Contingency Table 4.7 was related to the “B” passing grade. The fact that 40% of the students in G3 and one-third (34%) of students in G2 had GPAs below 2.00 at the end of the freshman year was surprising. Those students were failing more often than even the “C” students in G1. Moreover, none of the G3 and only 3% of the G2 students were able to get GPAs above 3.00.

Further analysis of the freshman year GPAs revealed that the students who passed the BUEPT with “A” were notably more successful than “B” and “C” students. First

of all, the percentage of failures among “A” students was significantly lower than the percentage of failures among both “B” and “C” students (8% in G1, 20% in G2 and 17% in G3). Furthermore, 64% of G1, 35% of G2 and 70% of G3 students had GPAs above 2.50 at the end of the freshman year.

Summarizing the results of the above three Contingency Tables (Tables 4.5, 4.6 and 4.7) three important points should be mentioned. First, students in all three groups – G1, G2, G3 – had the highest number of failures at the end of the first semester and the lowest number of failures at the end of the second semester. Second, among the three groups G2 was the least successful and G1 was the most successful group throughout the whole freshman year. Third, attending preparatory classes for a longer period of time did not ensure a more successful academic life for the subjects of the current study.

4.3.2 Comparison of Means within Groups

Tables 4.8, 4.9 and 4.10 show the results of the One-way analysis of variance aiming to check whether there were any significant differences among the means of those students who passed the BUEPT with different grades (i.e., A, B or C). Tables 4.8 and 4.9 show the data for the first and second semester respectively, while Table 4.10 reveals the results for the end of the freshman year.

4.3.2.1 Multiple Comparison of the First Semester Means

One-way analysis of variance of the first semester means ended up with different results for each of the three groups (see Table 4.8).

In G1, the “A” Group ($M1A= 2.66$) had a mean that was significantly higher than the means of both the “B” ($M1B= 2.35$) and the “C” ($M1C= 2.28$) Groups. However, there was not a significant difference between the means of those groups who passed the BUEPT either with “B” ($M1B= 2.35$) or “C” ($M1C= 2.28$).

Table 4. 8
Multiple Comparison – ANOVA- Scheffe – FIRST SEMESTER

		COMPARED GROUPS		MEAN DIFF.	STD. ERROR	SIG
FIRST TRY (G1)	Means	A	B	.3124*	.117	.038
	As = 2.66	A	C	.4174*	.124	.004
	Bs = 2.35	B	C	.1053	.115	.66
	Cs = 2.28					
1SEM PREP (G2)	Means	A	B	.3125	.185	.246
	As = 2,27	A	C	.5437*	.179	.013
	Bs = 1,96	B	C	.2312	.156	.337
	Cs = 1,73					
2SEMS PREP (G3)	Means	A	B	.3513*	.137	.041
	As = 2,39	A	C	.7285*	.138	.000
	Bs = 2,04	B	C	.3772*	.138	.026
	Cs = 1,67					

*. The mean difference is significant at the .05 level.

In G2, there was a significant difference only between the means of those who passed the BUEPT either with “A” ($M2A= 2.27$) or “C” ($M2C= 1.73$). As Table 4.8 exhibits, no significant difference was found either between the means of those students who passed the exam with “A” ($M2A= 2.27$) or “B” ($M2B= 1.93$), or between those who passed it with “B” ($M2B= 1.93$) or “C” ($M2C= 1.73$).

In G3, however, all of the differences among the examined means were significant (M3A= 2.39; M3B= 2.04; M3C= 1.67). Thus, it seems that if a student has had two semesters of prep, his/her academic success in the freshman year first semester of study at BU can easily be predicted by looking at his/her BUEPT grade.

4.3.2.2 Multiple Comparison of the Second Semester Means

Table 4. 9
Multiple Comparison – ANOVA- Scheffe – SECOND SEMESTER

		COMPARED GROUPS		MEAN DIFF.	STD. ERROR	SIG
FIRST TRY (G1)	Means	A	B	.2927	.142	.122
	As = 2.62	A	C	.2635	.123	.105
	Bs = 2.32	B	C	.0292	.115	.968
	Cs = 2.35					
1SEM PREP (G2)	Means	A	B	.3239	.181	.208
	As = 2.62	A	C	.7162*	.176	.000
	Bs = 2.29	B	C	.3923*	.153	.041
	Cs = 1.90					
2SEMS PREP (G3)	Means	A	B	.4310*	.120	.002
	As = 2.50	A	C	.4548*	.121	.001
	Bs = 2.07	B	C	.0270	.121	.975
	Cs = 2.04					

*. The mean difference is significant at the .05 level.

As Table 4.9 displays, in G1, none of the differences between the second semester means of the “A”, “B” and “C” students were found to be significant (M1A= 2.62; M1B= 2.32; M1C= 2.35). Moreover, the mean of the “Cs” was higher than the mean of “Bs”. That is, the BUEPT passing grade did not show an effect on the second semester academic achievement of the students in G1.

When the means of the “A”, “B” and “C” students in G2 were compared, significant differences were found between the means of both “As” (M2A= 2.62) and “Bs”

($M2B = 2.29$) and, "As" ($M2A = 2.62$) and "Cs" ($M2C = 1.90$). No significant difference was found between the means of "A" and "B" students.

In contrast to the first semester, at the end of the second semester no significant difference was found between the means of "B" ($M3B = 2.07$) and "C" ($M3C = 2.04$) students in G3. However, the differences between the means of those who passed the exam with "A" ($M3A = 2.50$) and "B" ($M3B = 2.07$), and "A" ($M3A = 2.50$) and "C" ($M3C = 2.04$) were still significant.

4.2.2.3 Multiple Comparison of the Freshman Year Means

Table 4. 10
Multiple Comparison – ANOVA- Scheffe – END OF THE FRESHMAN YEAR

		COMPARED GROUPS		MEAN DIFF.	STD. ERROR	SIG
FIRST TRY (G1)	Means	A	B	.3121	.143	.094
	As = 2.70	A	C	.3400*	.102	.004
	Bs = 2.39	B	C	.0375	.095	.924
	Cs = 2.32					
1SEM PREP (G2)	Means	A	B	.3182	.154	.124
	As = 2.44	A	C	.6226*	.149	.000
	Bs = 2.13	B	C	.3044	.129	.069
	Cs = 1.82					
2SEMS PREP (G3)	Means	A	B	.3912*	.104	.001
	As = 2.44	A	C	.5933*	.104	.000
	Bs = 2.06	B	C	.2021	.104	.157
	Cs = 1.85					

*The mean difference is significant at the .05 level.

As Table 4.10 shows, in G1, no significant difference was found between the means of those students who passed the exam with "A" ($M1A = 2.7$) or "B" ($M1B = 2.39$), and between those who passed it with "B" ($M1B = 2.39$) or "C" ($M1C = 2.32$). A

significant difference was found only between the means of those groups of students who passed BUEPT either with "A" ($M_{1A}=2.7$) or "C" ($M_{1C}=2.32$) ($p<.05$).

Similar results were obtained for G2 as well. Except for the significant difference found between the means of those students who passed the exam with "A" ($M_{2A}=2.44$) or "C" ($M_{2C}=1.82$), no significant difference was found between other means ($M_{2A}=2.44$ vs. $M_{2B}=2.13$; $M_{2B}=2.13$ vs. $M_{2C}=1.82$).

In contrast to the results obtained for the first two groups, in G3, an analysis of variance showed that there were significant differences between the means of groups who passed BUEPT with "A" ($M_{3A}=2.44$) and "B" ($M_{3B}=2.06$), and the groups who passed with "A" ($M_{3A}=2.44$) and "C" ($M_{3C}=1.85$). No significant difference was found between the means of those who passed the exam with either "B" or "C".

PART B: The Answer to Research Question 3

To obtain a clear picture of what students think of BUEPT a questionnaire was administered to all subjects.

4.4 Students' Questionnaire Answers

In this section the answers of students to each question in the second part of the Students' Questionnaire will be presented and discussed.

4.4.1 Analysis of the First Question

Table 4.11
Parts of BUEPT according to their difficulty level

	First Try (G 1)	One Sem. Prep (G2)	Two Sems. Prep (G3)	Total (G)
Listening	1.32	1.45	1.02	1.26
Reading	2.93	2.67	2.84	2.81
Writing	2.20	2.12	2.14	2.15

1 = the easiest, 3 = the most difficult

When the subjects were asked to order the parts of BUEPT (listening, reading, and writing), depending on their level of difficulty, they all agreed on the same order. According to the students, reading is the most difficult component of BUEPT ($M_R=2.81$), followed by writing ($M_W=2.15$) and listening ($M_L=1.26$).

4.4.2 Analysis of the Second Question

The second question in the second part of the questionnaire concerned the ideas of students about the main goal(s) of preparatory classes. Students were given a chance to choose more than one option (see Appendix H). The reason for this kind of application was to prevent respondents from thinking that there is only one major aim of prep classes, which might in turn lead to not being able to obtain the real thoughts and feelings of the informants (Nunan, 1992).

Moreover, in order to see whether attending prep classes affects the perspective from which the students evaluate the test or not, both those who passed the exam at first try and those who spent a semester or two in YADYOK were asked to answer this question.

Table 4.12
What is the main goal of prep classes?

		1	2	3	4	5	6	7	8
		To teach students how to pass BUEPT	To prepare students for their future academic success	To develop students' reading skill	To develop students' listening skill	To develop students' writing skill	To develop students' speaking skill	To improve the level of English proficiency of students	Other
First Try (G1)	N	167	28	98	81	89	23	35	22
	%	32	5	18	15	16	4	6	4
One Sem Prep (G2)	N	81	18	48	32	35	12	25	16
	%	31	7	18	12	13	4	9	6
Two Sems Prep (G3)	N	116	21	59	48	52	18	59	3
	%	30	6	15	13	14	5	16	1
Total (G)	N	364	67	205	161	176	53	119	41
	%	31	6	17	14	15	4	10	3

The analysis of data revealed that student answers are two-fold (see Table 4.12). First, the “preparation for the BUEPT” (25%), and second the “development of the three language skills” - reading (17%), writing (15%), and listening (14%). An interesting finding of the informants’ answers to this question was the fact that they did not include the development of the speaking component as a major goal of prep classes (4%). BUEPT does not have a speaking component, therefore, according to students, this skill is not developed in prep classes.

In addition to those two main aims presented above, informants also stated the following as possible main goals of preparatory classes: “to improve the level of English language proficiency of students” (10%), “to prepare students for their future academic studies” (6%) and “other” (3%).

When the answers of subjects in different groups (G1 vs. G2 vs. G3) were examined, it was found that the answers of students did not vary much across the groups (see Table 4.12). They all agreed on the idea that the major goal of prep classes was preparing students for BUEPT.

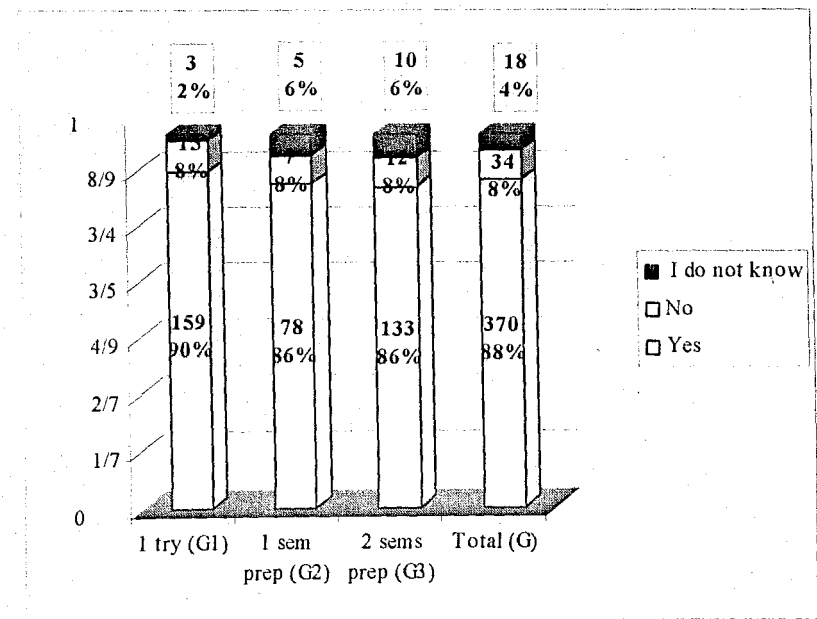
The findings in this section explain the results displayed in Table 4.1 (see Section 4.1). The reason why the students who attend prep classes for a longer period of time perform significantly better on BUEPT can be explained by the fact that YADYOK both directly (i.e., by teaching students test techniques) and indirectly (i.e., by developing skills that are measured on the exam) aims only at preparing students for the BUEPT. This claim is supported by students’ statements that teachers, especially in advanced classes, tended to ignore activities and topics which

did not contribute directly to passing the BUEPT. Informants put down: “All the compositions we wrote throughout the year required us to compare and contrast two things. Aren’t there any other types of compositions?” or “Teachers always were beginning their sentences with the same words: On the Proficiency exam...”

It seems as if the curriculum and instruction of a whole institution – YADYOK – is arranged around the BUEPT.

4.4.3 Analysis of the Third Question

Figure 4.2
Should there be an exam measuring your level of academic English proficiency before you begin your freshman year studies?
Distribution of students’ answers according to the time they spent in YADYOK



When asked whether there should be an exam measuring their level of academic English proficiency or not, the majority of the students (88%) said “Yes”, 8% said

“No” and only 4% chose the “I do not know” option (see Figure 4.2). The examination of the answers of each of the three groups to this question revealed that there was an agreement among them. Ninety percents of the subjects in G1, and 86% of the subjects in G2 and G3 believed that there should be an exam measuring the level of academic English proficiency of incoming students.

Table 4. 13
Should there be an exam measuring your level of the academic English proficiency before you begin your freshman year studies?

WHY DO YOU THINK SO?

YES = 370 = 88%	
Category	Number
1. ACADEMIC SUCCESS	193
2. REPUTATION OF BU	74
3. MOTIVATION	44
4. ADAPTATION/ORIENTATION	40
5. NO ANSWER	19
NO = 34 = 8%	
Category	Number
1. BUEPT	25
A: Does not measure what it claims to measure	10
B: Students' academic life should not depend only on BUEPT	5
C: English Proficiency Test is needed, but BUEPT with its current form and content should not be used	10
2. AUTONOMY: Entering the exam should be voluntary	8
3. NO ANSWER	1
I DO NOT KNOW = 18 = 4%	
Category	Number
1. AUTONOMY	7
2. NO ANSWER	11

When asked to explain their answers, most of the subjects (193 out of 370) who chose “Yes” stated: “It is really difficult, if not impossible, to have a successful academic life in FLED, LL, and TRANS Departments of BU without a high level of English proficiency” (see Table 4.13). A second group of informants (74 out of 370) said that BU is reputable in terms of the level of English of its graduates. In order to

keep those high standards, no students with low English proficiency should be allowed to start their studies at the university.

The following two groups respectively indicated that such a proficiency exam would motivate them to improve their English language proficiency and language skills (44 out of 370), and would help them to adjust to the atmosphere of BU (40 out of 370). Nineteen out of 70 students did not give any explanation to why they said “Yes”.

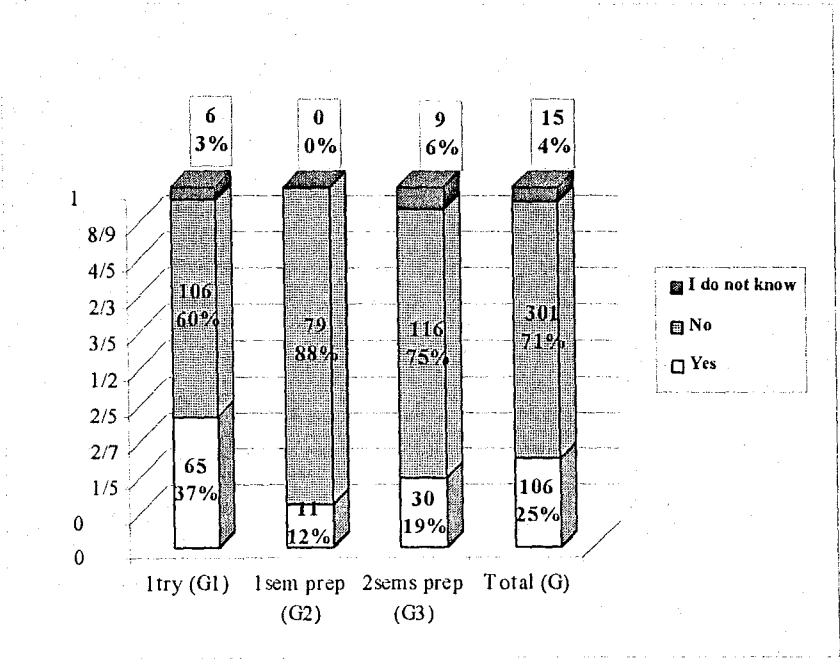
On the other hand, the explanations of the informants in relation to their opinion about not having an English language proficiency test can be gathered around three major categories: “BUEPT”, “autonomy”, and “no answer” (see Table 4.13).

The majority of the informants (25 out 34) stated their opinions in relation to the first category – BUEPT - which was criticized from three different perspectives: “BUEPT does not measure what it claims to measure”; “the future academic life of students should not depend only on BUEPT”; and “an English proficiency test is needed but BUEPT with its current form and content should not be used”. A second small group of respondents (8 out of 34) expressed their belief that at university level, some of the decisions related to their academic life should be left to students. According to those respondents, students themselves, not an English proficiency test, should decide whether their level is adequate or not for academic studies. One student did not write any explanation about why there should not be an English proficiency test at the beginning of academic studies.

Of the 18 informants who chose the “I do not know” option, 11 wrote no explanations for their response while 7 declared that “taking the exam should be voluntary”.

4.4.4 Analysis of the Fourth Question

Figure 4. 3
Did BUEPT accurately reflect your level of academic English proficiency?
Distribution of students’ answers according to the time they spent in YADYOK



The fourth question in the second part of the questionnaire requested students to indicate whether the BUEPT accurately reflected the level of their academic English proficiency or not. As Figure 4.3 displays, 25% of students said “Yes”, 71% said “No” and only 4% ticked “I do not know”. On examination of the answers of the different groups, it can be seen that the majority of the students in each group

responded negatively to this question (60% in G1, 88% in G2, and 75% in G3). However, it should be pointed out that those who did not attend preparatory classes (G1) gave a higher number of positive answers (37%), than those who spent a semester or two in YADYOK (12% in G2, 19% in G3).

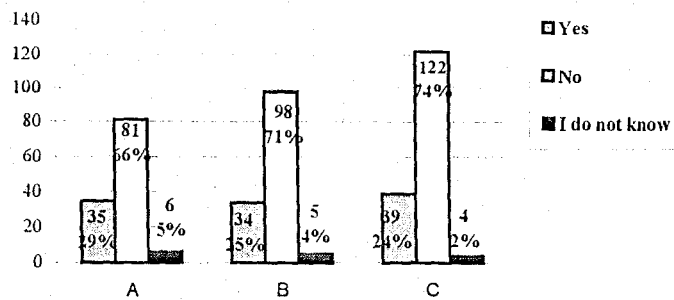
Results of Figure 4.3 can be explained with the expressions “irregular” for G2 and “longer university life” for G3. Students in G2 were not allowed to start their undergraduate studies at the beginning of the academic year because they had failed the BUEPT at first try. After a semester of prep, G2 students passed the exam, but this time they had to take the undergraduate courses in the “wrong” order. In other words, they have to take the second semester courses before taking the first semester courses. This, in itself, made G2 students’ academic life harder. Not surprisingly (see Table 4.2, Section 4.2.1), at the end of the freshman year, G2 was the group with the lowest academic achievement.

On the other hand, G3 was the group that passed the exam at third try. That means an extra year at university and extra financial load for the students’ families. According to students, the sole reason behind this prolongation was BUEPT, which did not fulfill its proposed role, i.e., to measure their level of English language proficiency.

The examination of the distribution of students’ answers according to their BUEPT passing grade - A, B or C - revealed no significant difference among the three groups (see Figure 4.4). The majority of the students in each group (66% of As, 71% of Bs

and 74% of Cs) chose the “No” option and around one-fourth of each group ticked “Yes” (29% of As, 25% of Bs and 24% of Cs). Briefly, passing BUEPT with different grades – A, B or C – did not affect subjects’ attitudes towards the test.

Figure 4.4
Did BUEPT accurately reflect your level of academic English proficiency?
Distribution of students’ answers according to their BUEPT passing grade



When the answers to the “Why do you think so?” part of the question were scrutinized (see Appendix H), it was noticed that the positive replies tended to fall into four main categories: “skills”, “selection”, “parallelism”, “no answer” (see Table 4.14).

One-third (33 out of 106) of the students explained their “Yes” answer with the fact that “BUEPT measures the three important skills – reading, writing and speaking – and their sub skills that have a great importance for the success in academic life.” The other one-third of students (33 out of 106) reported that BUEPT was “really a difficult, selective and challenging test” which measured their level of academic English proficiency better than the ÖYS English Test. Twenty students affirmed that

BUEPT measured “exactly the level of their English proficiency”. Still another twenty students (out of 106) provided no reason for their affirmative answer.

Table 4. 14
Did BUEPT accurately reflect your level of academic English proficiency?
WHY DO YOU THINK SO?

YES = 106 = 25%	
Category	Number
1. SKILLS: measures skills important for a successful academic life	33
2. SELECTION: difficult and challenging; better than OYS English test	33
3. PARALLELISM: between BUEPT grades and students' self-evaluation	20
4. NO ANSWER	20
NO = 301 = 71%	
Category	Number
1. BUEPT	189
A: Aim of the test: measures test techniques and test tricks	113
B: Parallel test forms: quality and difficulty level of BUEPT change	54
C: Authenticity: very structured; does not allow to show the real English proficiency level; not related to the requirements of first year courses	22
2. ADMINISTRATION	112
A: Listening: success on this part depends on the reader	49
B: Organizational problems: students in different classes do not do the same question at the same time	37
C: Time Limits: exam is too long and tiring	26
I DO NOT KNOW = 15 = 4%	
Category	Number
NO ANSWER	15

On the other hand, the analysis of the negative explanations ended up with two different categories: 1) “BUEPT” and 2) “Test administration” (see Table 4.14).

1) BUEPT

More than half of the negative explanations were directly related to the BUEPT itself (189 out of 301). Subjects listed three different reasons – aim, parallelism and

authenticity - why they thought the exam was not an effective measure of their level of academic English proficiency.

According to the first and the largest group of students (113 out of 176), BUEPT did not measure accurately their level of English proficiency since that was not its intention. Those respondents pointed out that the exam aimed and measured their knowledge of test techniques. Nearly all of them wrote: "If you know the test techniques, you can pass BUEPT even if your English is not good enough".

Some techniques often listed by students were:

"Just combine the prompts given in the writing part with "hence", "however", "nevertheless" and "thus". Do not bother about the meaning and content.", or

"Do not read the text in the reading part. Just try to find the sentences which have the same vocabulary as the question.", or

"Start taking notes only when names, reasons, levels, or parts of a process are listed.

That is, after phrases such as: "The following four reasons are stated" or "And finally, I want to list the names of the most famous writers." "

Another group of students (54 out of 189) stated that the level of difficulty and the quality of the BUEPT change from one time to another (i.e., problems involving test reliability). According to them, whether or not the testees pass the exam depends mostly on two things. First, on the exam itself and second, on the time exam is taken. One of these students wrote:

“The first time I took the exam, it was really easy but I failed since I did not know the techniques for answering the listening and reading questions. The second BUEPT that I took was an exam completely different from the first one. Both the reading passages and the writing topics were really very difficult, as if they had been prepared so that students would not be able to answer them.” Another explanation written by the students in this group was: “Sometimes it’s as if the test writers want all students who take the exam to pass it, and sometimes as if they want everybody to fail.”

Other twenty-two students related their “No” answer to the authenticity of the exam. Informants declared that BUEPT is a very structured exam that does not allow students to show their real level of English. Moreover, as believed by subjects in this group, the things required on the exam are artificial and different from the things required in their departments.

2) Administration

One third of the subjects (112 out of 301) pointed out that for them BUEPT was not a good tool for measuring their level of academic English proficiency since there were serious problems with the administration of the test (see Table 4.14).

Among the criticisms related to the test administration, the higher percentage is related to the problem of confirmity of test conditions (49 out of 112). As was mentioned before, due to the low quality of the available equipment, the ‘talk’ in the

listening part is delivered live (see Section 2.1.2, Listening). That is, a number of trained 'readers' go from class to class during the test and read the lectures to the students (although it is clear that there are possible implications for the reliability of the test). However, a large number of informants stated that different readers deliver the talks in different ways. For example, "while some read quickly, others wait for students to note everything"; or "while some just read smoothly the text, others stress some parts which is a clue for testees that this is the answer to the question". Stated differently, most of the students share the belief that whether you get a passing grade on the listening section depends in large part on the person who delivers the speech. Students summarized the problem with the following sentences: "Sometimes it is very difficult to pass the BUEPT. It is not because you do not know English or test techniques but because you are not lucky enough to be in a class where the reader of the listening part reads slowly and has an understandable accent."

Table 4.14 shows that another very important criticism related with the administration of BUEPT is the order in which the different parts of the test are administered (37 out of 112). It is known that one of the basic principles of test preparation and administration is to order and administer different parts of the test from the easiest to the most difficult one (Harrison, 1983; Heaton, 1988; Weir, 1995; Brown, 1996). As the discussions in Section 4.4.1 revealed, according to the subjects of the current study, reading is the most difficult and listening is the easiest component of BUEPT. Nevertheless, due to the lack of enough trained 'readers', some classes have the listening part at the beginning of the exam, while others have it at the end, and still others have it in the middle. In other words, different classes do not take the different parts of the exam in the same order, which is "a serious

reliability issue” (Brown, 1996; Anastasi & Urbina, 1997), that might, in turn, affect the validity of BUEPT.

This fact, as many of the informants stated, puts the students who have to take the reading part at the very beginning in a disadvantaged position compared to the other two groups. A student wrote on her questionnaire: “If you have to do the reading part first and then in the listening part you have a teacher who reads quickly and with an accent you are not accustomed to, then forget passing BUEPT.”

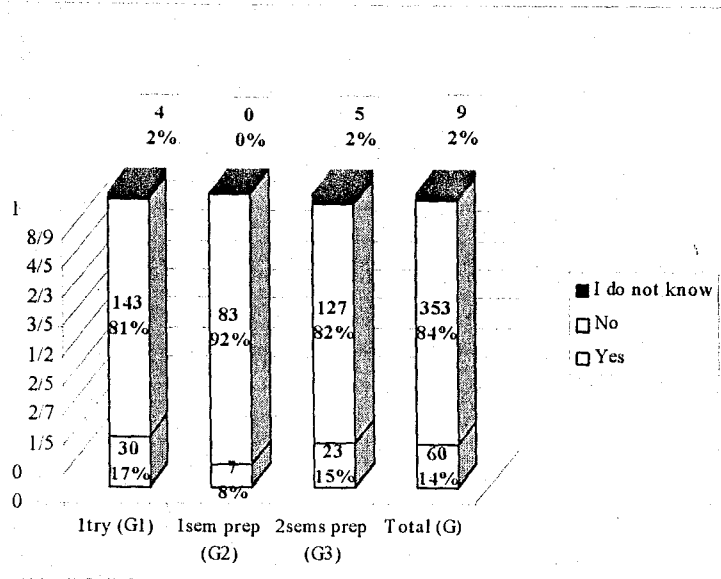
Many of the students who took the reading part first admitted that after seeing the questions asked in that section they became discouraged and decided that their English was not good enough for university study and that they had to attend prep classes. Because of this, they did not do the following two parts.

The last problem stated by students (26 out of 112) in relation to the test administration was timing. Students in this group think that BUEPT does not accurately measure their level of academic English proficiency because “the exam is very long and exhausting”. They mentioned that there are too many questions and that they are too difficult and most of the students do not have “enough energy to do the questions in the last section”. That is why, according to examinees, there is “a serious discrepancy between the grade on the exam and the real level of English proficiency of students”. None of the students who chose the “I do not know” option as an answer to the fourth question of the questionnaire explained their answers.

4.4.5 Analysis of the Fifth Question

Figure 4.5
Did BUEPT properly predict your freshman year academic success?

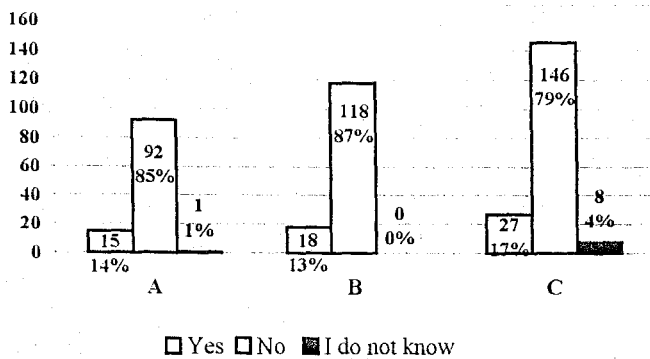
Distribution of students' answers according to the time they spent in YADYOK



When asked whether BUEPT properly predicted their freshman year academic success nearly all students said “No” (84%), 14% said “Yes” and only 2% said “I do not know” (see Figure 4.5). While analyzing the responses in different groups, it was noticed that the rates of answers are pretty similar to those given to the previous questions (see Figures 4.2 and 4.3). G2 is the group with the highest rate of negative answers (92%), G1 is the group with the lowest rate of negative answers (81%) and G3 is between them with 82% of negative answers.

Figure 4.6

Did BUEPT properly predict your freshman year academic success?
Distribution of students' scores according to their BUEPT passing grade



Further analysis intending to investigate whether or not the BUEPT passing grade affects examinees' thoughts and attitudes towards the exam was carried out (see Figure 4.6). The analysis ended up with finding a parallelism between the answers of the subjects in the three groups. Most of the students in "A" (85%), "B" (87%) and "C" (79%) groups stated that BUEPT did not predict their freshman year academic success. Interestingly, more students among those who passed the exam with C (17%) stated that the proficiency exam was a good predictor of their future academic attainment than among those who passed with A (14%) or B (13%).

The examination of the students' explanations revealed that the small number of students who gave affirmative answers to the fifth question in the second part of the questionnaire, mainly justified their responses by stating that there was a parallelism between their BUEPT grades and their freshman year GPAs (see Table 4.15). The rest of the students in "Yes" group (18 out of 60) supplied no explanations for their answers.

Table 4.15
Did BUEPT properly predict your freshman year academic success?
WHY DO YOU THINK SO?

YES = 60 = 14%	
<u>Category</u>	<u>Number</u>
1. PARALLELISM: Between BUEPT grades and the final grades of the freshman year courses	42
2. NO ANSWER	18
NO = 353 = 84%	
<u>Category</u>	<u>Number</u>
1. DOUBLE NEGATIVE: BUEPT had not measured properly the level of academic English proficiency of students, thus it could not predict their academic success	201
2. LACK OF PARALLELISM: between BUEPT grades and the freshman year GPA	121
3. NO RELATION: academic success in BU is not related to the level of English language proficiency of students	17
4. NO ANSWER	14
I DO NOT KNOW = 9 = 2%	
<u>Category</u>	<u>Number</u>
1. NO RELATION: academic success in BU is not related to the level of English language proficiency of students	2
2. NO ANSWER	7

As can be seen from Table 4.15, negative explanations can be gathered under four headings: 1) Double negative; 2) Lack of parallelism; 3) Lack of relationship; 4) No answer.

Nearly two-thirds (201 out of 353) of the subjects in the “No” group argued that the BUEPT could not predict their freshman year academic success simply because it did not measure their level of academic English proficiency. According to them, the test techniques required by the exam did not help them and were not related to the things required in academic life.

The second group of students (121 out of 353) explained their negative answers with the fact that their freshman year performance was different from their BUEPT

passing grade. While some of the informants wrote: "Although I passed BUEPT with A, I had a hard time in our department". Others pointed out that even though they got C on the exam, they had a pretty successful academic life. For example a student said: "Passing with C did not discourage me, at the end of the freshman year, my GPA was nearly four out of four."

Still another small group of subjects (41 out of 353) claimed that BUEPT could not predict their academic attainment because academic success in BU is not related to the level of English language proficiency but to other factors such as "effort, intelligence and good relationships with professors". The last twenty-five students in the "No" group did not provide any reasons for their negative answers.

Of the nine students who chose the "I do not know" option, two argued that academic success was not related to the level of English language proficiency, while seven wrote no explanations.

CHAPTER 5

DISCUSSION AND CONCLUSIONS

The starting point of this study was to evaluate the English Language Proficiency Test – BUEPT - used at Boğaziçi University (BU). The aim of this study was to find out whether the test really differentiated between students who would face difficulty because of poor language abilities or study skills from those who would not. Since the scores obtained on BUEPT were considered to have serious consequences on the future academic life of students at BU, the study also aimed to obtain and analyze qualitative feedback from students about the operation and adequacy of BUEPT. In this section, the major findings of this research will be summarized and the implications of the results will be discussed.

The first finding of this study (see Table 4.1) supports Henning's (1990:380) concern about the effect of testing on teaching. He says that: "If there is no concerted effort to subordinate testing to explicit curricular goals, there is an ever-present potential danger that tests themselves with all their inherent limitations will become the purpose of the educational encounter by default." This was also the case detected in the current study. YADYOK, as observed by Onat (1996:27) and confirmed by students (see Table 4.12), directly or indirectly, intends to prepare students only for the BUEPT. Prep students master skills and techniques that are measured on the exam. All other course content that is not included in the test is ignored. According to the subjects of the study, the development of the speaking skill, one which cannot be thought of as a separate part of the quartet of language skills, is not among the main goals of YADYOK. This claim is supported by the results of BÜMED's

(1996) study, where informants agreed that they were bad at speaking, while they were really good at the other three skills. The finding stated above may be accepted as a possible explanation for the large number of subjects in the BÜMED study (1996) who were not satisfied with their English education at YADYOK.

Additionally, from other findings of the study, it becomes clearer why students who attended prep classes performed significantly better on the BUEPT than those who passed the exam at first try (see Figure 4.1 and Table 4.1.). This finding is similar to Amer (1993) and, Robb and Ercanbrack's (1999) studies which illustrated that "teaching to the test" did in fact result in higher test scores.

According to Enginarlar (1984:129), "the success of any formal language program is crucially linked with its testing philosophy and practice". That is, whether or not the effort of a whole institution – YADYOK, which is organizing each of its activities around BUEPT, wastes its time or really does something valuable depends on the validity of the test which is applied.

BUEPT is a proficiency test, "in order to be valid it has to look forwards" (Alderson, 1984:33), it has to be concerned with the future success of the examinees. Nevertheless, negative relationships were found to exist between the BUEPT scores and the first semester, the second semester and the overall freshman year GPAs of the students in FLED, LL and TRANS Departments of BU. That is, BUEPT did not do its job. The majority of the students who were able to pass BUEPT with high grades (A or B) and were also expected to be successful in their academic studies (GPAs above 2.50) failed (i.e., had GPAs below 2.00), while some of the students

who could hardly pass the exam manifested a good academic performance during their freshman year.

This finding is consistent with the results of studies conducted by Sugimoto (1966), Mulligan (1966), Hwang and Dizney (1970), Sharon (1972), Gue and Holdaway (1973), Shay (1975), Wilcox (1975), Jensen (1980), Odunze, (1980), Hale et al., (1983), Traynor (1985), Zeidner (1986, 1987), Light et al. (1987), Vinke and Jochems (1993), Qualls and Ansley (1995), Tonkyn (1995) and Yan (1995), which detected that different language proficiency tests used as screening and placement procedures in different institutions failed to predict the academic performance of the incoming students of those institutions. Results also show that although the main aim of proficiency tests is to test and predict the future performance of examinees (Davies, 1988; Heaton, 1988; Hughes, 1989; Bachman, 1990; Brown, 1996), this is rarely achieved.

On the other hand, the results of this study did not confirm the findings of Burgess and Greis (1970), Heil and Aleamoni (1974), Ayers and Peters (1977), Baldauf and Dawson (1980), Freidenburg and Curry (1981), Rea (1984), Ho and Spinks (1985), and Dedo (1990), which detected a statistically significant correlation between English language proficiency test scores and GPAs.

When the first semester, the second semester and the overall freshman year GPAs of each of the three groups – G1, G2, G3 - were examined, it was detected that the most successful group on the exam was not also the academically most successful group (see Table 4.2 and Table 4.3). That is, G3, the group that performed significantly better on BLUEPT, could not attain GPAs that were significantly higher than the

GPA's of the other two groups in the two semesters of the freshman year. On the other hand, G1, the group with the lowest achievement level on BUEPT, was the most successful group during the freshman year.

The general information presented in the three Contingency Tables (see Tables 4.5, 4.6, and 4.7) can be summarized as follows. First, students in all three groups – G1, G2 and G3 – had the highest number of failures at the end of the first semester and the lowest number of failures at the end of the second semester. Second, the percentage of failures among the students in G2 and G3 (i.e., those students who passed the BUEPT after a period of prep) was significantly higher than the percentage of failures in G1 throughout the whole freshman year. Third, the differences between the academic performances of the three groups were most obvious at the end of the first semester, nevertheless these differences were minimized at the end of the second semester (e.g., Means of the groups at the end of the first semester $M1_{SEM1}=2.39$, $M2_{SEM1}=1.93$, $M3_{SEM1}=2.04$; Means of the groups at the end of the second semester $M1_{SEM2}=2.40$, $M2_{SEM2}=2.19$, $M3_{SEM2}=2.20$).

Data in the Contingency Tables (see Tables 4.5, 4.6, and 4.7) also displayed three very important facts about the relationship between the BUEPT passing grade and the first semester, the second semester and the overall freshman year GPA's. Firstly, in each of the three groups, those students who passed the BUEPT with "A" had a better academic attainment than those who passed the exam with "B" or "C" during the freshman year. However, "B" students were not always (e.g., especially during the second semester) more successful than "C" students. Secondly, students who passed the exam after a semester or two of prep had more failures even than the "C"

students in G1. Besides that, “C” students in G2 and G3 had three times more failures at the end of the first semesters (70%) and two times more failures at the end of the second semester (42%) and the freshman year (60,5%) than the students who passed the exam at first try (27%; 21% and 28%). Considering all these findings, it can be concluded that Doltaş and Sevgen’s (1995) claim that the determining factor which indicates whether students will pass the BUEPT or not is not their level of English but their familiarity with the test and test techniques, is correct.

Originally, Hughes (1988:137) intended to prepare a test that was “best suited to the development of the English skills necessary for study at Boğaziçi University”. Nevertheless, one of the conclusions that may be drawn from this research is that the BUEPT did not, in fact, really measure the language skills that subjects of this study needed in order to be academically successful. This conclusion was based on two findings. First, all of the correlations between the BUEPT passing grades and the first semester, the second semester and the overall freshman year GPAs of the three groups are negative. And second, as the period of time spent in YADYOK increased, the predictive validity of BUEPT decreased. That is, BUEPT predicted the freshman year success of G1 the best (though still negatively) and the academic performance of G3 the worst. However, G3 was the group that attended the preparatory classes the longest period of time, as compared to G1 and G2. On the exam, as it was expected, G3 was the most successful group. If the exam had really measured the skills necessary for academic achievement at BU, then the relation between the GPAs of the students in G3 and their BUEPT scores would have been the strongest. In fact, this was not the case. The strongest relationship were found between the BUEPT scores and the first semester, the second semester and the

overall freshman year GPAs of G1 students, who were not exposed to the English education at YADYOK.

This finding backs up Robinson and Ross's (1996) claim that skill-focused tests only provide an indication of a student's current knowledge of the English language and cannot be used to predict the actual academic skills of students. The solution to the problems which were encountered through the use of skill-based tests came from Paltridge (1992) and, Robinson and Ross (1996). They designed integrated-approach tests that were more successful in predicting academic success of students than the traditional skill-based tests.

Another point that should be discussed in relation to BUEPT is its cut off point which is 60 out of 100. Data presented in the Contingency Tables (see Tables 4.5, 4.6 and 4.7) did not support the argument that BUEPT intends "to distinguish those students whose English is adequate for study at the University through the medium of English from those whose English is not" (Hughes, 1984:137). When the freshman year academic performance of "C" students was scrutinized it was noticed that at the end of the first semester 27% of G1, 71% of G2 and 68% G3 students were among the failures. At the end of the second semester, the rates of students with GPAs below 2.00 in each of the groups were as follows 21% in G1, 53% in G2 and 31% in G3. At the end of the freshman year, one-third of the students in G2 and two-thirds of the students in G2 and G3 (60% in G2 and 61% in G3) became "repeaters". It appears that BUEPT seriously overpredicted the freshman year academic achievement of those who passed it with "C"; and it was especially

unsuccessful in predicting the first semester academic attainment of “C” students in G2 and G3.

Thus, it can be concluded that this study provided no empirical support for the BUEPT cut off point. Not having “empirical underpinning” is considered one of the most important failings of proficiency tests by measurement experts (Lantolf & Frawley, 1985, 1988; Fulcher, 1996).

These findings are similar to the results of Light et al.’s (1987), Vinke and Jochems’s (1993) and Yan’s (1995) studies which revealed that the cut off points of TOEFL used by different institutions were not successfully set. Moreover, the results of the present study back up Zeidner’s (1986, 1987) conclusion that language proficiency tests sometimes tend to mispredict the future attainment of the test-takers.

In his study Fulcher (1997:118) used Popham’s “pooled judgements” technique to set the cut off points of tertiary level proficiency tests. It is possible that this method may be used to determine the cut off point of the BUEPT as well. This would need further study and analysis by testing experts in YADYOK.

The results obtained from the in-groups ANOVA analysis provided no empirical rationale for the ranges with which different grades – A, B, C - were counterparted. As Tables 4.8, 4.9 and 4.10 indicated, significant differences were found only between the means of those students who passed the BUEPT either with “A” or “C”. However, generally in all groups, the GPAs of those who passed the exam with “A” were not found to be significantly higher than the GPAs of those who passed it with

"B". In addition, the GPAs of those who passed the BUEPT with "B" were not found to be significantly higher than the GPAs of those who passed it with "C". These findings support Doltaş and Sevgen's (1995) claim that the A, B, C ranges of BUEPT are not sensitive to the different levels of language proficiency of students. They went even further by saying that the ranges were chosen as if at random. Moreover, these findings imply that there should not be a grade like "B" and the BUEPT grading scheme should be revised on the basis of a pass/fail system.

The findings mentioned above supported the argument of Gue and Holdaway (1973), Traynor (1985), Saville-Troike (1991), Qualls (1995), Tonkyn (1995), and Burston et al. (1996) who claimed that the selection procedure of students who would be allowed to continue their academic studies at tertiary level, should not be based solely on their proficiency in English as measured by different English proficiency tests.

In answer to the third research question of the study it was found that students neither perceived the exam as a valid reflector of their level of academic English proficiency nor as a good predictor of their future academic performance. Similar to previous studies (Low, 1982; Wall et al., 1994) there was a close parallelism between the results of the statistical analysis and the students' evaluations of the test. These results implied that there were really some problems related to the test, and the test takers were aware of these problems. The findings also seem to back up Mendeloshn's (1989) statement that students had very clearly defined opinions about the tests they were given.

Students also voiced some very important reliability issues related to the preparation (e.g., lack of parallelism between the BUEPTs administered at different times) and administration (i.e., different “readers”; different groups of students taking the three parts of BUEPT in different order) of BUEPT. According to Anastasi and Urbina (1997:13), “any influence that is specific to a test situation constitutes error variance and reduces test validity”. Maybe these were among the reasons why most of the subjects in the present study also saw the BUEPT as “a barrier rather than a bridge to educational opportunity” (Tonkyn, 1995:37).

On the other hand, the eagerness of all the students to participate in this study, the large amount of data supplied by them on the questionnaire supported Frisbie (1982), Low (1985) and Fulcher’s (1997) claims that obtaining qualitative feedback from students on the operation of a test is very important.

Students’ responses to the third question in the second part of the questionnaire supported Dietel, Herman, and Knuth’s (1991) claim about the existence of a continuous faith in the value of assessment for stimulating and supporting “school improvement and instructional reform at national, state, and local levels”. Although subjects of the study had negative feelings and attitudes towards BUEPT, nearly all of them (88%) stated that there should be an exam measuring their level of academic English proficiency. They also expressed their beliefs that exams motivated and guided them towards future academic achievement.

The findings of the three research questions posed in this study and the discussions of these findings illustrate that proficiency tests like BUEPT should be systematically revised and adjusted according to the needs of the test-takers. As Frisbie (1982:140) puts it, tests are "... subject to malfunction over time and must be revised as curricula and students change". Since 1983 many things have changed. Technological developments and tools such as computers and Internet have become a part of students' everyday lives. Also, the expectations and the characteristics of students accepted to BU have changed. On the other hand, as it is well known, language methodology is not static. It is continuously being revised and modified in keeping with the most up-to date ideas on second language acquisition. As a result of this, the courses in different departments of BU and their contents change. Only in the Foreign Language Education Department, to give an example, the undergraduate program and courses have changed four times since 1983. Despite all these changes, BUEPT with its format, content, and language testing techniques has been the same throughout all these years. That is why it is now time for some innovations.

Since a test cannot be valid without being reliable (and BUEPT has some very important reliability problems stated by students) the first thing that YADYOK test-designers and administrators should do is to improve the reliability of BUEPT. The available test examples should be thoroughly examined to check whether the "parallel forms of BUEPT" are really parallel. In addition, the organizational climate of BUEPT should be improved. That is, the administration procedures should be reorganized and standardized.

Furthermore, another reliability issue has to do with the fact that BUEPT is being administered over years to both undergraduate and (unfortunately) graduate students. This is certainly inimical to Hughes's original intentional design of the test, which was based on a needs analysis of freshman year undergraduate students.

Another problem that relates to needs analysis has to do with updating the test. As the needs of various departments change overtime based on social and cultural conditions, so should the composition of tasks and skills required by BUEPT. Therefore, YADYOK teachers and instructors from different faculties should cooperate to improve the authenticity of BUEPT regarding the changing reality of BU.

The third alternative that might help YADYOK administrators to gain time is to utilize widely used standardized tests such as the new computer-based TOEFL as a criterion. That is, students might be required to take this test in addition to taking BUEPT. A later investigation on the predictive validity of the standardized test used may reveal whether or not BU really needs a "special" test for itself. If it proves to be useful in differentiating students with limited English proficiency for academic studies from those with a sufficient level of English proficiency, a widely used standardized test can be employed instead of BUEPT. This would then be a good standing point for the standardization of the requirements of the BU in accordance with universities all over the world. This may in turn lead university administrators to think that the level of English proficiency of students in BU is the same as the proficiency level of students in other reputable universities all over the world. If this cannot be done for financial reasons, it can be suggested to the testing experts at

YADYOK that they utilize the results and standardization procedures of the standardized tests in finding a way to revise and standardize the BUEPT.

In making inferences from the present study, some limitations need to be considered.

The first limitation arose with respect to the subjects of the study. In the current research only with an English major were considered. Some of the results may well reveal the specific views of these students, and may not reflect the attitudes and thoughts of informants from other departments. It would seem advisable to replicate this study among students from different departments in BU, in order to determine the extent to which the results obtained can be generalized. It can also be suggested that the study be repeated for different academic years to check if the same results are still obtained.

Secondly, correlating the overall BUEPT scores only with the overall (first semester, second semester and freshman year) GPAs offered very limited information i.e., it did not reflect many aspects of the students' academic involvement that might affect the final outcome. It would be interesting to investigate the relationships between the sub-tests scores of BUEPT and the final grades of the different freshman year courses.

Thirdly, it is also necessary to gather evidence about the internal validity and reliability of BUEPT (i.e., sub-tests intercorrelations, test homogeneity, inter-marker

correlations, or item analysis). This analysis might end up with possible explanations for the high negative correlations between the scores on the BUEPT and the students' academic success.

As a concluding remark it can be said that further research including the judgements of the YADYOK teachers and the assessments of the freshman year instructors, besides the GPAs of the students, would be useful in order to find even more precise results and generalize the findings of the present study. It is also hoped that this study will inspire other researchers to do further work in the area of language testing.

REFERENCES

- Abadzi, H.** (1975). Evaluation of foreign students admission procedures used at the University of Alabama. *Dissertation Abstracts International*, 7754-A.
- Alderson, J. C.** (1978). The affect of certain methodological variables on cloze test performance and its implication for the use of the cloze procedure in EFL testing. Fifth International Congress of Applied Linguistics, Monreal.
- Alderson, J. C.** (1981). Report of the discussion on communicative language testing. In Alderson, J. C. & Hughes, A. (Eds.), *Issues in Language Testing* (pp. 55-65). ELT Document 111, London: The British Council.
- Alderson, J. C.** (1984). Testing English for specific purposes: How specific can we get? In Bozok, S. & Hughes, A. (Eds.), *Proceeding of the Seminar, Testing English Beyond the High School* (pp. 33-51). Istanbul: Boğaziçi University Publications.
- Amer, A. A.** (1993). Teaching EFL students to use test-taking strategy. *Language Testing*, 10 (1), 71-77.
- Anastasi, A.** (1954). *Psychological Testing (First Edition)*. New York: Macmillan.
- Anastasi, A. & Urbina, S.** (1997). *Psychological Testing (Seventh Edition)*. Upper Saddle River, New Jersey: Prentice Hall.
- Andalib, A. A.** (1975). The academic success of undergraduate Iranian students in selected Texas University. *Dissertation Abstracts International*, 4881-A.
- Angoff, W. H.** (1988). Validity: An evolving concept. In Wainer, H. & Braun, H. (Eds.), *Test Validity* (pp. 19-32). Hillsdale: Erlbaum.
- Ascher, C.** (1990). Assessing bilingual students for placement and instruction. (ERIC Document Reproduction Service No. ED322273).

- Ayers, J. B. & Peters, R. M. (1977). Predictive validity of the test of English as a foreign language for Asian graduate students in engineering, chemistry, or mathematics. *Educational and Psychological Measurement*, 37, 461-463.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly*, 25 (4), 671-704.
- Baldauf, R. B. Jr. & Dawson, R. L. T. (1980). The predictive validity of the Michigan test of English language proficiency for teacher trainees in Papua New Guinea. *Educational and Psychological Measurement*, 40, 1201-1205.
- Barber, E. G., Morgan, R. P. & Torstrick, R. L. (1987). Foreign graduate students in U.S. engineering programs: Problems and solutions. *Engineering Education*, 78, 171-174.
- Boğaziçi University. (1996). *Boğaziçi University Catalogue 1996-1998*. Istanbul: Boğaziçi University Publications.
- Boğaziçi University. (1997). *Boğaziçi Üniversitesi İngilizce Yeterlilik Sınavı*. Istanbul: Boğaziçi University Publications (see Appendix A).
- Boğaziçi University. (1998). *Boğaziçi Üniversitesi El Kitabı 1998*. Istanbul: Boğaziçi University Publications.
- Bostic, M. L. (1981). A correlational study of academic achievement and the test of English as a second language (TOEFL). *Dissertation Abstracts International*, 42, 468A.
- Bozok, S. & Hughes, A. (Eds.) (1984). *Proceeding of the Seminar, Testing English Beyond the High School*. Istanbul: Boğaziçi University Publications.

- Brindley, G.** (1986). *The Assessment of Second Language Proficiency: Issues and Approaches*. Sydney: National Curriculum Resource Centre.
- Brown, J. D.** (1996). *Testing in Language Programs*. New Jersey: Prentice Hall.
- Burgess, M. L. & Greis, N. B.** (1970). English language proficiency and academic achievement among students of English as a second language at the college level. (ERIC Document Reproduction Service No. ED 074812).
- Burston, J. L., Monville-Burston, M. & Harfouch, J.** (1996). Comparison of evaluation measures in the placement of first-year students. *Occasional Papers - Applied Linguistics of Association of Australia*, 15, 59-78.
- BÜMED.** (1996). Boğaziçi Üniversitesi Performans Araştırması. İstanbul: BÜMED.
- Canale, M.** (1983). On some dimensions of language proficiency. In Oller, J. W. (Ed.), *Issues in Language Testing Research* (pp. 333-342). Rowley, Mass.: Newbury House.
- Canale, M. & Swain, M.** (1980). Theoretical basis of the communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Carroll, J. B.** (1961). Fundamental considerations in testing for English language proficiency of foreign students. In *Testing English Proficiency of Foreign Students* (pp. 30-40). Washington, DC: Center for Applied Linguistics.
- Carroll, J. B.** (1980). *Testing Communicative Performance: An Interim Study*. Oxford: Pergamon.
- Cattell, R. B.** (1964). Validity and reliability: A proposed more basic set of concepts. *Journal of Educational Psychology*, 55, 1-20.

- Chaplen, E. F.** (1970). The Identification of Non-Native Speakers of English Likely to Under-Achieve in University Courses Through Inadequate Command of the Language. Unpublished Ph.D. Thesis. University of Manchester.
- Chomsky, N.** (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.
- Clark, J. L. D.** (1972). *Foreign Language Testing: Theory and Practice*. Philadelphia, Pa.: Center for Curriculum Development, Inc.
- Clark, J. L. D.** (1976). *Direct testing of Speaking Proficiency: Theory and Practice*. Princeton, New Jersey: Educational Testing Service.
- Cohen, L. & Holliday, M.** (1982). *Statistics for Social Science*. London: Harper & Row.
- Corbin, S. S. & Chiachiere, F. J.** (1995). Validity and reliability of a scale measuring attitudes toward foreign language. *Educational and Psychological Measurement*, 55 (2), 258-267.
- Couchene, R. J. & de Bagheera, J. I.** (1985). A theoretical framework for the development of performance tests. In Hauptman, P. C., Le Blanc, R. & Wesche, M. B. (Eds.), *Second Language Performance Tests* (pp. 45-58). Ottawa: University of Ottawa Press.
- Cronbach, L. J.** (1949). *Essentials of Psychological Testing*. New York: Harper & Row.
- Cronbach, L. J.** (1971). Validity. In Thorndike, R. L. (Ed.), *Educational Measurement* (pp. 443-597). Washington, DC: American Council on Education.

- Cronbach, L. J.** (1984). *Essentials of Psychological Testing (Fourth Edition)*. New York: Harper & Row.
- Davies, A.** (1978a). Language testing. Survey article Part I. *Language Teaching and Linguistics Abstracts*, 11 (3), 145-159.
- Davies, A.** (1978b). Language testing. Survey article Part II. *Language Teaching and Linguistics Abstracts*, 11 (4), 215-231.
- Davies, A.** (1988). Procedures in language test validation. In Hughes, A. (Ed.), *Testing English for University Study* (pp. 29-35). ELT Documents 127. Oxford: Modern English Press.
- Dedo, D. R.** (1990). Profile of the ESL freshman writer: Predicting success in the academic setting. *Dissertation Abstracts International*, 52 (3), 889A.
- Del Vecchio, A. & Guerrero, M.** (1995). Handbook of English language proficiency tests.
<http://www.ncbe.gwu.edu/miscpubs/eacwest/elpteststhtml>
- Dietel, R. J., Herman, J. L. & Knuth, R. A.** (1991). What does research say about assessment? NCREL, Oak Brook.
http://www.ncrel.org/sdrs/areas/stw_esys/4assess.ht
- Doltaş, D. & Sevgen, C.** (1995). Institutional TOEFL ve B.Ü. Proficiency Test sonuçlarının karşılaştırılması. BU Unpublished Research Project of Western Languages and Literatures Department.
- Douglas, D.** (1986). From school to university: Language policy and performance at the University of Khartoum. *International Journal of the Sociology of the Language*, 61, 98-112.

- Douglas, D. & Chapelle, C. (Eds.)** (1993). *A new Decade of Language Testing Research: Selected Papers from the 1990 Language Testing Research Colloquium*. Virginia: Teachers of English to Speakers of Other Languages, Inc.
- Duran, R. P., Canale, M., Penfield, J., Stansfield, C. W., & Liskin-Gasparro, J. E.** (1987). TOEFL from a communicative viewpoint on language proficiency: A working paper. In Freedle, R. O., Duran, R. P. (Eds.), *Cognitive and Linguistic Analyses of Test Performance* (pp. 1-150). Norwood, New Jersey: Ablex.
- Enginarlar, H.** (1984). The problems of large scale testing: The METU case. In Bozok, S. & Hughes, A. (Eds.), *Proceedings of the Seminar, Testing English Beyond the High School* (pp. 115-131). Istanbul: Boğaziçi University Publications.
- Farhady, H.** (1982). Measures of language proficiency from the learners' perspective. *TESOL Quarterly*, 16 (1), 31-58.
- Frederiksen, J. R. & Collins, A.** (1989). A system approach to educational testing. *Educational Research*, 18 (9), 27-32.
- Freidenburg, J. E. & Curry, W. C.** (1981). English proficiency and the bilingual university student. (ERIC Document Reproduction Service No. ED 208657).
- Frisbie, D. A.** (1982). Methods of evaluating course placement systems. *Educational Evaluation and Policy Analysis*, 4 (2), 133-140.
- Fulcher, G.** (1996). Invalidating validity claims for the ACTFL oral rating scale. *System*, 24 (2), 163-172.
- Fulcher, G.** (1997). An English language placement test: Issues in reliability and validity. *Language Testing*, 14 (2), 113-138.

- Gradman, H. L. & Reed, D. J.** (1997). Assessment and second language teaching. In Bardovi-Halig, K. & Hartford, B. (Eds.), *Beyond Method Components of Second Language Teacher Education* (pp.199-213). New York: McGraw-Hill Book Company.
- Graham, J. G.** (1987). English language proficiency and the prediction of the academic success. *TESOL Quarterly*, 23 (2), 505-521.
- Gray, B. T.** (1997). Controversies regarding the nature of score validity: still crazy after all these years. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Austin.
- Gue, L. R. & Holdway, E. A.** (1973). English proficiency tests as predictors of success in graduate studies in education. *Language Learning*, 23 (1), 89-103.
- Guilford, J. P.** (1946). New standards for test evaluation. *Educational and Psychological Measurement*, 6, 427-439.
- Guilford, J. P. & Fruchter, B.** (1973). *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill Book Company.
- Gulliksen, H.** (1950). Intrinsic validity. *American Psychologist*, 5, 511-517.
- Hale, G., Stansfield, C. & Duran, R.** (1983). *Summaries of TOEFL Studies, 1963-1982*. (Research Report No.16). Princeton, NJ: Educational Testing Service.
- Harmer, J.** (1993). *The Practice of English Language Teaching*. London: Longman
- Harris, D. P.** (1969). *Testing English as a Second Language*. New York: McGraw-Hill Book Company.
- Harris, D. P. & Palmer, L. A.** (1970). *Comprehensive English Language Test for Speakers of English as a Second Language*. New York: McGraw-Hill.

Harrison, A. (1983). *A Language Testing Handbook*. London: Macmillan.

Harrison, A. (1990). *A Language Testing Handbook*. London: Macmillan.

Heaton, J. B. (1988). *Writing English Language Tests*. London: Longman.

Heil, D. K. & Aleamoni, L. M. (1974). Assessment of the proficiency in the use and understanding of English by foreign students as measured by the Test of English as a Foreign Language. (ERIC Document Reproduction Service No. ED 093948).

Heikinheimo, P. S. & Shute, J. C. M. (1986). The adaptation of foreign students: Student views and institutional implications. *Journal of College Student Personnel*, 27, 399-406.

Henning, G. (1990). Priority issues in the assessment of communicative language abilities. *Foreign Language Annals*, 23 (5), 379-384.

Hernandez-Chavez, E., Burt, M. & Dulay, H. (1978). Language dominance and proficiency testing: Some general considerations. *NABE Journal*, 3, 41-45.

Hinofotis, F. B. (1981). Perspectives of language testing: Past, present and future. *Nagoya Gakuin Daigaku Gaikokugo Kyoiku Kiyo*, 4, 51-59.

Ho, D. Y. F. & Spinks, J. A. (1985). Multivariate prediction of academic performance by Hong Kong University students. *Contemporary Educational Psychology*, 10, 249-259.

Hubley, A. M. & Zumbo, B. D. (1996). A dialectic on validity: Where we have been and where we are going. *The Journal of General Psychology*, 123 (3), 207-215.

- Hughes, A.** (1984). The new Boğaziçi University English Language Proficiency Test. In Bozok, S. & Hughes, A. (Eds.), *Proceedings of the Seminar, Testing English Beyond the High School* (pp. 137-147). Istanbul: Boğaziçi University Publications.
- Hughes, A.** (1988). *Testing English for University Study*. ELT Documents 127. Oxford: Modern English Press.
- Hughes, A.** (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hughes, A. & Porter, (Eds.)** (1983). *Current Developments in Language Testing*. London: Academic Press.
- Hwang, K. Y. & Dizney, H. F.** (1970). Predictive validity of the test of English as a foreign language for Chinese graduate students at an American University. *Educational and Psychological Measurement*, 30, 475-477.
- Hymes, D. H.** (1967). Models of interaction of language and social setting. *Journal of Social Issues*, 33, 8-28.
- Hymes, D. H.** (1972). On communicative competence. In Pride, J. B. & Holmes, J. (Eds.), *Sociolinguistics*. Harmondsworth: Penguin Books.
- Jones, R. L.** (1977). Testing: A vital connection. In Phillips, J. K. (Ed.), *The Language Connection: From the Class to the World* (pp. 237-265). Skokie, Ill.: National Textbook Company.
- Jensen, A. R.** (1980). *Bias in Mental Testing*. New York: Free Press.
- Jonson, J. L. & Plake, B. S.** (1998). A historical comparison of validity standards and validity practices. *Educational and Psychological Measurement*, 5 (58), 736-753.

- Kinnell, M. (Ed.) (1990). *The Learning Experiences of Overseas Students*. Buckingham: Society for Research into Higher Education and Open University Press.
- Lado, R. (1961). *Language Testing*. New York: McGraw-Hill Book Company.
- Lantolf, J. P. & Frawley, W. (1985). Oral proficiency testing: A critical analysis. *The Modern Language Journal*, 69, 337-45.
- Lantolf, J. P. & Frawley, W. (1988). Proficiency: Understanding the construct. *Studies in Second Language Acquisition*, 10 (2), 181-195.
- Light, R. L., Xu, M. & Mossop, J. (1987). English proficiency and academic performance of international students. *TESOL Quarterly*, 21 (2), 251-261.
- Low, G. (1982). Report on the use of student feedback in a testing program. *University of Hong Kong, Language Centre Working Papers in Linguistics and Language Teaching*, 5, 52-54.
- Low, G. D. (1985). Validity and the problem of direct language proficiency test. In Alderson, C. (Ed.), *Evaluation* (pp. 151-168). Lancaster: Pergamon Press.
- Low, G. D. (1986). Storylines and other developing contexts in use-of-language test design. *Indian Journal of Applied Linguistics* (Special Issue on Language testing).
- McMillan, J. H. & Schumacher, S. (1989). *Research in Education*. London: Scott Foresman and Company.
- Mendelsohn, D. J. (1989). Testing should reflect teaching. *TESOL Canada Journal*, 7 (1), 95-108.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012-1027.

- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In Wainer, H & Braun, H. (Eds.), *Test Validity* (pp. 33-45). Hillsdale, New Jersey: Erlbaum.
- Messick, S. (1989). Validity. In Linn, R. L. (Ed.), *Educational Measurement (Third Edition)* (pp. 13-103). London: Macmillan.
- Messick, S. (1995). Validity of psychological assessment. Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50 (9), 741-749.
- Mestre, J. P. (1981). Predicting academic achievement among bilingual Hispanic college technical students. *Educational and Psychological Measurement*, 41, 1255-1264.
- Miller, G. A. & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behaviour*, 2, 217-228.
- Morrison, T. & Morrison, M. (1995). A meta-analytic assessment of the predictive validity of the quantitative and verbal components of the graduate record examination with graduate point average representing the criterion of graduate success. *Educational and Psychological Measurement*, 55 (2), 309-316.
- Morrow, K. E. (1977). *Techniques of Evaluation for a Notional Syllabus*. London: Royal Society of Arts.
- Morrow, K. E. (1979). Communicative language testing: Revolution or evolution. In Brumfit, C. J. & Johnson, K. (Eds.), *The Communicative Approach to Language Teaching* (pp. 143-158). Oxford: Oxford University Press.
- Mosier, C. L. (1947). A critical examination of the concept of face validity. *Educational and Psychological Measurement*, 7, 191-205.

- Mostiçyan, S. (1979). English language background and academic success: A survey of the first year students at Boğaziçi University. Unpublished M.A. Thesis, Department of Guidance and Psychological Counselling. Boğaziçi University
- Mulligan, A. C. (1966). Evaluating foreign credentials. *College and University*, 41, 307-313.
- Nunan, D. (1992). *Research Methods in Language Learning*. Cambridge: Cambridge University Press.
- Odunze, O. J. (1980). Test of English as a foreign language and first year GPA of Nigerian students. *Dissertation Abstracts International*, 42 (8), 3419-3420A.
- Oller, J. W. Jr. (1976). Evidence for a general language proficiency factor: An Expectancy grammar. *Die Neueren Sprachen*, 75, 165-174.
- Oller, J. W. Jr. (1978). Pragmatics and language testing. In Spolsky, B. (Ed.), *Advances in Language Testing, Series 2*. Arlington, VA: Centre for Applied Linguistics.
- Oller, J. W. Jr. (1979). *Language Tests at School*. London: Longman.
- Oller, J. W. Jr. (1981). Language testing research 1979-1980. In Jones, R. L. & Tucker, G. R. (Eds.), *Annual Review of Applied Linguistics: 1980*. Rowley, Mass.: Newbury House.
- Oller, J. W. Jr. (Ed.) (1983). *Issues in Language Testing Research*. Rowley, Mass.: Newbury House.
- Oller, J. W. Jr. (1984). 'g', What is it? In Hughes, A. & Porter, D. (Eds.), *Current Developments in Language Testing* (pp. 35-37). London: Academic Press.

- Oller, J. W. Jr., & Damico, J. S. (1991). Theoretical considerations in the assessment of LEP students. In Hamayan, E. & Damico, J. S., *Limiting Bias in the Assessment of Bilingual Students*. Austin: Pro-Ed Publications.
- Oller, J. W. Jr. & Perkins, K. (Eds.) (1978). *Language in Education: Testing the Tests*. Rowley, Mass: Newbury House.
- Onat, Z. (1997). The Use of Referential Questions in the EFL Classroom. Unpublished M.A. Thesis. Department of Foreign Language Education, Boğaziçi University.
- Paltridge, B. (1992). EAP Placement testing: An integrated approach. *English for Specific Purposes*, 11, 243-268.
- Perkins, K. (1980). Using objective methods of attained writing proficiency to discriminate among holistic evaluations. *TESOL Quarterly*, 14 (1), 61-69.
- Pilliner, A. E. G. (1968). Subjective and objective testing. In Davis, A. (Ed.), *Language Testing Symposium: A Psycholinguistic Approach* (pp. 19-35). London: Oxford University Press.
- Qualls, A. L. & Ansley, T. N. (1995). The predictive relationship of ITBS and ITED as measures of academic success. *Educational and Psychological Measurement*, 55 (3), 485-498.
- Rea, P. M. (1978). Assessing language as Communication. *MALS Journal*, 3, 5-13.
- Rea, P. M. (1984). Language tests as indicators of academic achievement. (ERIC Document Reproduction Service No. ED 275172).
- Read, J. A. S. (Ed.) (1981). Papers on Language Testing. *SEAMEO Regional Language Centre*, Occasional Papers No. 18.

- Robb, T. N. & Ercanbrack, J. (1999). A study of the effect of direct test preparation on the TOEIC scores of Japanese university students. *TESL-EJ*, 3 (4), A-2.
<http://www.kyoto-su.ac.jp/information/tesl-ej/ej12/a2.html>
- Robinson, P. & Ross, S. (1996). The development of task-based assessment in English for academic purposes programs. *Applied Linguistics*, 17 (4), 455-474.
- Savignon, S. J. (1972). Teaching for communicative competence: a research report. *Audio Visual Language Journal*, 10 (3), 153-162.
- Saville-Troike, M. (1991). Teaching and testing for academic achievement: the role of language development. *NCBE FOCUS: Occasional Papers in Bilingual Education*, No. 4.
<http://www.ncbe.gwu.edu/ncbepubs/focus/focus4.html>
- Schachter, J. (1990). Communicative competence revised. In Harley, B., Allen, P., Cummins, J. & Swain, M. (Eds.), *The Development of Second Language Proficiency* (pp. 39-49). Cambridge: Cambridge University Press.
- Scott, M. L. (1986). Student affective reactions to oral language tests. *Language Testing*, 3 (1), 99-118.
- Sharon, A. T. (1972). English proficiency, verbal aptitude, and foreign students success in American graduate schools. *Educational and Psychological Measurement*, 32, 425-431.
- Shay, H. R. (1975). Affect of foreign students' language proficiency on academic performance. *Dissertation Abstracts International*, 36, 1983A.
- Skehan, P. (1988). State of the art article. Language testing Parts I and II. *Language Teaching*, 21 (4), 235-242.

- Slark, J. & Bateman, H. (1982). A study of non-native English speakers' academic performance. *Dissertation Abstracts International*, 36, 1983A.
- Spolsky, B. (1968). Preliminary studies in the development of techniques for testing overall second language proficiency. Problems in foreign language testing. *Language Learning*, 3, 39-49.
- Spolsky, B. (1971). Reduced redundancy as a language testing tool. In Perren, G. E. & Trim, J. L. (Eds.), *Application of Linguistics*. Cambridge: Cambridge University Press.
- Spolsky, B. (Ed.) (1978). *Approaches in Language testing*. Advances in Language Testing, Series: 2. Arlington, Va.: Center for Applied Linguistics.
- Spolsky, B. (1985). What does it mean to know how to use a language? An essay on the theoretical basis of language testing. *Language Testing*, 2 (2), 180-191.
- Spolsky, B. (1989). *Conditions for second Language Learning*. Oxford: Oxford University Press
- Stevenson, D. (1981). Beyond faith and face validity: The multitrait-multimethod matrix and the convergent and discriminat validity of oral proficiency tests. In Palmer, A. S., Groot, P. J. M. and Trosper, G. A. (Eds.), *The Construct Validation of Tests of Communicative Competence*. Washington, DC: TESOL.
- Stevenson, D. (1982). Pop validity and performance testing. Paper presented at the International Symposium on Language Testing, University of Hong Kong.
- Stevenson, D. (1985). Authenticity, validity, and a tea party. *Language Testing*, 2 (1), 41-47.
- Stover, A. D. (1982). Effects of language admission criteria on academic performance of non-native English-speaking students. *Dissertation Abstracts International*, 36, 1983A.

- Sugimoto, R. A.** (1966). The relationship of selected predictive variables to foreign students' achievement at the University of California, Los Angeles. *Dissertation Abstracts International*, 28, 65A.
- Taylor, D. S.** (1988). The meaning and use of the term 'competence' in linguistics and Applied Linguistics. *Applied Linguistics*, 9, 14-68.
- Tonkyn, A.** (1995). English language proficiency standards for overseas students: Who needs what level? *International Education*, 6, 37-61.
- Traynor, R.** (1985). The TOEFL: An appraisal. *ELT Journal*, 39 (1), 43-47.
- Valette, R. M.** (1977). *Modern Language Testing*. New York: Harcourt Brace Jovanovich, Inc.
- Vinke, A. A. & Jochems, W. M. G.** (1993). English proficiency and academic success in international postgraduate education. *Higher Education*, 26, 275-285.
- Wall, D., Clapham, C. & Alderson, J. C.** (1994). Evaluating a placement test. *Language Testing*, 11 (3), 321-344.
- Weir, C. J.** (1988). The specification, realisation and validation of an English language proficiency test. In Hughes, A. (Ed.), *Testing English for University Study* (pp. 45-110). ELT Documents 127, Oxford: Modern English Press.
- Weir, C. J.** (1990). *Communicative Language Testing*. London: Prentice Hall.
- Weir, C. J.** (1995). *Understanding and Developing Language Tests*. London: Phoenix ELT.

- Wen, Q. & Johnson, R. K. (1997). L2 learner variables and English achievement: A study of tertiary-level English majors in China. *Applied Linguistics*, 18 (1), 27-48.
- Widdowson, H. G. (1983). *Learning Purpose and Language Use*. Oxford: Oxford University Press.
- Wilcox, L. O. (1975). The prediction of academic success of undergraduate students. *Dissertation Abstracts International*, 35, 6084B.
- Yan, Y. (1995). The TOEFL paradigm and its current issue: a further study of predictability of academic success for international graduate students at Mississippi State University. *Dissertation Abstracts International*, 56 (6), 2213-A.
- Zeidner, M. (1986). An English language aptitude tests biased towards culturally different minority groups? Some Israeli findings. *Language Testing*, 3 (1), 80-98.
- Zeidner, M. (1987). A comparison of ethnic, sex and age bias in the predictive validity of English language aptitude tests: some Israeli data. *Language Testing*, 4 (1), 55-71.

APPENDIX A



BOĞAZIÇI ÜNİVERSİTESİ İNGİLİZCE YETERLİLİK SINAVI
BOĞAZIÇI UNIVERSITY ENGLISH PROFICIENCY TEST

-BUEPT-

A. SINAVLA İLGİLİ GENEL BİLGİLER

SINAVIN BÖLÜMLERİ

1. Duyduğunu Anlama Bölümü (60 dakika)

Adaylar iki konuşma dinlerler. Birinci konuşmada, adaylar konuşma başlamadan önce soruları okurlar ve konuşmayı dinlerken bu soruları cevaplandırır. İkinci konuşma sırasında ise adaylar not tutarlar ve ancak konuşma bittikten sonra, soruları okuyup tuttukları notlara dayanarak cevaplandırır.

2. Okuduğunu Anlama Bölümü (60 dakika)

Adaylara iki ayrı okuma metni verilir. Adaylar bu metinlerle ilgili soruları cevaplandırır.

Birinci metinle ilgili soruları cevaplandırmak için, metnin hızlı bir şekilde taranması gerekmektedir. İkinci metin ile ilgili sorular ise daha dikkatli bir okuyup, anlama gerektirir.

3. Yazılı İfade (80 dakika)

Bu bölümde adaylardan, herbiri 40'ar dakikalık iki kompozisyon yazmaları istenir. Adaylar yazacakları konularda not şeklinde küçük bilgilerle yönlendirilir.

SINAVIN BÖLÜMLERİNİN AĞIRLIĞI

Sınavın bölümlerinin ağırlıkları aşağıdaki şekilde belirtilmiştir:

Duyduğunu Anlama Bölümü	% 30
Okuduğunu Anlama Bölümü	% 40
Yazılı İfade	% 30

DEĞERLENDİRME SİSTEMİ

Sınavın değerlendirmesi, aşağıda gösterilen not baremine göre yapılır:

A |—————→

B |————→

C |————→

BAŞARILI

F1 |————→

F2 |————→

F3 |————→

BAŞARISIZ

Sınavın "Duyduğunu Anlama ve Okuduğunu Anlama" bölümleri 6 ayrı kış; Yazılı İfade bölümü 4 ayrı kış tarafından değerlendirilir.

B. İNGİLİZCE YETERLİLİK SINAVINDAN ÖRNEKLER

SAMPLE QUESTIONS FROM THE ENGLISH PROFICIENCY TEST

In order to help candidates prepare for the test, the following examples of test questions are given. The passages given as examples are much shorter than the ones which form part of the test itself.

LISTENNING COMPREHENSION

As noted earlier, this section is comprised of two talks. For the first talk, the candidate sees the questions in advance.

For example:

An Experiment in Psychology

Answer the following questions and fill in the blanks.

It is not necessary to answer in complete sentences.

1) When was the experiment carried out

2) a. What was the height of the larger block?

b. What was the weight of the smaller block?

3) What was the occupation of the subjects in the experiment?

4) On the average the block was perceived as being times as heavy as the other. In some cases it was perceived as being as much as times as heavy.

(and so on... In the actual test, there would be about 12 questions)

Candidates would then hear the talk, during which they would try to answer the questions.

"I am going to tell you about an experiment in psychology. It concerns the perception of weight. In this experiment, which was carried out in 1954, the experimenter prepared two rectangular blocks of wood. The size of these blocks is important. One was 4 cm. by 4 cm. by 6 cm.; the other was 8 cm. by 8 cm. by 12 cm. - so this block was exactly eight times the volume of the first. Inside these blocks the experimenter had put pieces of lead. These pieces of lead were adjusted in such a way that the total weight of each block was 300 gm. Each of the two blocks had a metal ring attached at the top. A block could be lifted by placing the forefinger inside this ring.

Now for the experiment... The subjects in the experiment were 100 military officers. Each subject was told that the larger block weighed 300 gm. He was asked first the larger block... and then the smaller block. After he had done this he was asked to judge the weight of the smaller block in grams.

The results were surprising. Just about all of the subjects perceived the smaller block as being substantially heavier than the larger one. In fact the average estimate for the smaller block was 750 gm. - that is, 2.5 times the correct weight (which, you'll remember, was 300 gm.). For some of the subjects the overestimate was as much as sevenfold.

This phenomenon is known as the size-weight illusion. How is this phenomenon to be explained? Many people think that it happens because the person expects the **smaller** block to be lighter than the larger one, just because it is smaller. But when he lifts it, the contrast between what he expects and the actual weight of the smaller block makes it seem heavier.

But this doesn't really explain what happens, for the following four reasons. The **first** reason is this: the same person can go on lifting the two blocks repeatedly, and he will continue to say that the smaller one feels heavier. The **second** reason is that even after the subject has **weighed** both of the blocks on a **scale**, when he lifts them again, he will still say that the smaller one feels heavier. Knowledge of the true weights does not destroy the illusion. The **third** reason; the illusion is just as great when subjects are told to pay no attention to the relative size of the blocks.

And finally, the **fourth** reason; if the subject first looks at the blocks and then closes his eyes when he lifts them, the illusion is equally great.

Another explanation has been suggested. And it is this. Although the subject is attempting to judge **weight**, he is in fact implicitly judging **density**-density being weight per cubic centimetre. The small block is $1/8$ size of the large block, but weighs exactly the same-so the density of the small block is eight times that of the larger. So, perhaps when the subject is attempting to judge weight, he is affected by his perception of density. There is some experimental support for this explanation. If subjects are told to pay attention to the attribute of 'weight' and to ignore the attribute of 'density' then the size-weight illusion is reduced.

In the second part of the listening comprehension component, the candidates take notes while listening to another talk. They do not see the questions until the talk is **ended**. This second talk would normally be on quite a different topic from the first. For purpose of illustration, however, we will give examples of the kinds of questions that would have been based on the above talk, if it had been delivered in the second part of the listening component.

After taking notes, the candidate would be faced with questions of the following kind (for illustration, based only on the second half of the talk). It is not necessary to answer in complete sentences, or to write the exact words used in the talk.

1) What name has been given to the phenomenon described in the talk?

.....

2) A common explanation of the phenomenon is that the subject expects the smaller block to be than the larger. The between what he expects and the actual weight when he lifts it makes it seem

3) But this seems an unsatisfactory explanation because the effect persists when subjects:

a)

b)

c) are told to pay no attention to

d) look at the blocks, then while lifting the blocks.

4) An alternative explanation is that the subjects are in fact implicitly judging rather than weight.

(and so on... in the test itself, there would be about 12 questions.)

READING COMPREHENSION

Note: The following passage is shorter than those usually found in the test.

"Any discussion of movies and modern technology is certain to focus upon Charlie Chaplin and the problems he encountered as a factory worker in the film *Modern Times* (1936). American cinematic reaction to technology, however, began prior to Chaplin's portrayal of a beleaguered factory hand and it can be fully understood only if we consider the mechanical nature of the film itself, and the motion picture comic tradition that preceded *Modern Times*.

Modern Times, as well as any other film, comic or tragic, silent or sound, live or animated, must be considered in relation to one of the more significant developments of 19th century culture: the mechanization of the arts. Beginning with the invention of photography in the 1840's technology posed a threat to traditional artistic practices and forms, while at the same time it raised the possibility of novel techniques and wholly new art forms.

Arguments about the artistic status of photography and the role of the photograph as a source of inspiration for the painter or sculptor has not subsided when the phonograph appeared in the 1870's. Although the phonograph did not immediately raise questions about the creation of a new art form, as did the camera, it could be cited with photography as proof that artistic works, no matter how subtle or sublime, were amenable to **mechanical reproduction**. The advent of the camera and phonograph proved that art could be multiplied by a machine and prepared for wide distribution to the masses.

The culmination of the movement towards the mechanization of the arts was reached with the invention of motion pictures in the final decades of the 19th century. In cinema mechanical civilization found its natural mode of artistic expression: it was an art form was in essence mechanical, industrial, and commercial - a fact which is reflected in the characterization of early film as "canned" drama or machine-made theater. Among other things, the motion picture is a modern industrial product that results from the cooperative efforts of a group of artists, entrepreneurs, and technicians. To extend the industrial metaphor, it has been said that these makers of movies assemble their products in a manner analogous to that used on an automobile assembly line. Prior to the development of the cinema, no art form had ever been so dependent upon the mechanical and commercial world for its full realization."

Questions

- 1) Name all of the inventions mentioned in the passage.
- 2) What is the process of film making likened to?
- 3) What does the word "it" in the final sentence of the second paragraph refer to?
- 4) For each of the following, find a single word in the passage which has the same meaning.
 - a. (paragraphs 2-3), arrival
 - b. (paragraphs 3-4), period of ten years

WRITING

The following is an example of the kind of composition to be written:

Discuss the following proposed measures intended to increase the number of foreign tourists coming to Turkey. Do not use the word "I" what you write. You should write about one page.

- (1) More/better advertising and/or information
(where? what form should it take?)
- (2) Improved facilities (hotels, transportation, communications, etc.)
- (3) Training of personnel (guides, hotel managers, etc.)

Music & Society

while-listening

Nowadays, we hear music almost everywhere, in taxis, supermarkets and elevators, for example. Historically music has played an important role in all societies.

In Western culture, music is usually defined as a pleasing and expressive combinations of sounds. It goes without saying that different cultures have diverse views concerning the musical quality of different sounds. Thus, simple tribal chants or a composition created by a computer may or may not be accepted as music by members of a given society. Muslims, for example, do not consider the chanting of the Koran to be music. So, the social context of sounds may determine whether or not they are regarded as music. Industrial noises, for instance, are not music except when presented as part of a concert of experimental music.

Opinions also differ as to the origins and spiritual value of music. In some African cultures music is seen as something uniquely human. Among some native Americans it is thought to have originated as a way for spirits to communicate. In Western culture music is regarded as inherently good, and any verbal statement that is welcome is said to be 'music to the ears'. In some Eastern cultures, it is of low value, associated with sin and evil, and attempts have been made to outlaw its practice.

Music has many uses, and in all societies certain events are inconceivable without it. In many ways music can be compared to language. Both music and language are self contained systems of communication that must be learned to be understood.

In the West and in the high cultures of Asia, it is possible to distinguish between three basic levels of music. First, art or classical music, which is composed and performed by trained professionals; second, folk music, which is shared by the rural population at large and transmitted orally; and, third, popular music, performed by professionals, disseminated through radio, television, tapes and CDs, and consumed by a mass audience.

All societies have vocal music; and with few exceptions, all have instruments. Among the simplest instruments are sticks that are struck together, rattles, and body parts used to produce sound, as in slapping the thighs and clapping. Such simple instruments are found in many tribal cultures. At the other extreme, there are highly complex instruments such as pianos and violins that are capable of extremely subtle expression. More recently computers, synthesizers and samplers have greatly extended the expressive possibilities of music.

The normal method of remembering a piece of music and transmitting it is oral, and most of the world's music is learned by hearing. Everywhere music is used to accompany other activities. It is, for example, universally associated with dance. Although words are not found in singing everywhere, the association of music and poetry is so close that language and music are widely believed to have had a common origin.

Music is a major component in religious services, secular rituals, theatre, and entertainment of all sorts. In many societies it is also an activity carried on for its own sake. In Western society in the late 20th century, music is usually heard at concerts, on the radio, or at home on tape or CD. In many societies music serves as the chief entertainment at royal courts. The most widespread use of music, however, is as a part of a religious ritual. In some tribal societies, for example, music serves as a special form of communication with supernatural beings. Another, less obvious function of music is social integration. For most social groups, music can serve as a powerful symbol. Some minorities, for instance American blacks, use music as a major symbol of group identity.

Music may serve as a symbol in other ways. It can represent non-musical ideas or events, and it can underscore ideas that are verbally presented in operas, films and television drama. It also symbolizes military, patriotic and funerary feelings and events. In a more general sense, music may express the central values of a society. In Western music, the interrelationship of conductor and the various members of an orchestra symbolizes the need for strong coordination among specialists in a modern society.

In most of the world's societies, musicianship requires talent, special knowledge, training and effort. Thus, a successful musical work or performance is difficult to achieve. In tribal societies, the musician is regarded as someone who has supernatural powers. In certain early societies in Europe and America, however, musicians were regarded as undesirable social deviants. This remains the case in some parts of the present-day Middle East.

Each culture has its own music, and the classical, folk and popular traditions of a region are usually closely related and easily recognised as part of one system. The peoples of the world can be grouped musically into several large areas, each with its characteristic musical dialect. In the 20th century, however, rapid travel and mass communication have led to the phenomenon of world music. Musicians can now combine any styles that appeal to them, from the most traditional to the most modern.

KEY

Music & Society

while-listening

1. In the West music is defined as a pleasing & expressive combination of sounds.
2. What is the reaction of Muslims to the chanting of the Koran?
they do not consider it to be music
3. When are industrial noises considered as music?
when presented as part of a concert (of experimental music)
4. How is music regarded in some Native American cultures?
as a way for spirits to communicate
5. What is music linked with in some Eastern cultures?
sin & evil
6. What system of communication is music considered to be similar to?
language
7. What are the three different types of music in the West and Asia?
① art or classical music ② folk music ③ popular music
8. Give one example of how parts of the body are used to produce music.
slapping the thighs / clapping
9. Name two instruments that have extended the expressive range of music.
synthesizers and samplers
10. What are the two forms of art that are often coupled with music?
dance and poetry
11. What is the function of music in some tribal societies?
communication with supernatural beings
12. What does music signify in some minority groups?
group identity

13. From a general perspective, what does music represent?

it expresses the central values of a society

14. How is the musician regarded in tribal societies?

someone who has supernatural powers

15. How are musicians looked upon in some parts of the Middle East?

social deviants

16. What new musical development has been made possible by rapid travel and mass communication?

world music

GENDER AND COMMUNICATION (NT)

In this lecture, I'll talk about a variable that affects human communication. That variable is gender. Gender is the social identity that men and women learn as they grow up in a culture. For example, boys learn to be "masculine" and girls learn to be "feminine" as they grow to be men and women. Researchers have shown that men and women communicate in quite different ways and in different amounts, depending on the situation the speakers find themselves in, and the reasons they're communicating with other people.

Many cultures actually encourage men and women to talk differently and in different amounts, and these patterns for communicating are learned when men and women are very young. Children learn how to talk to other children or adults, and how to have conversations, not only from their parents but also from their peers. In her best-selling book *You Just Don't Understand*, Deborah Tannen points out that, although American boys and girls often play together, they spend *most* of their time playing in same-sex groups. She also points out that boys and girls *do* play some games together, but their favorite games are very often quite different. Tannen and other researchers on this topic have found that young boys, say ages 8 through 12, tend to play outside the house rather than in the house, and they play in large groups that are hierarchically structured. The group of boys generally has a leader who tells the other boys what to do and how to do it. It is by giving orders and making the other boys play by the rules that boys achieve higher or more dominant status in their play group. Boys also achieve status by taking "center stage." They take center stage by talking a lot, they give orders and commands; they tell a lot of stories and jokes. They command attention by dominating conversations and by interrupting other boys who are speaking. The researchers also found that boys' games often have clear winners and losers and elaborate systems of rules.

Researchers, on the other hand, found that girls play different kinds of games and abide by different rules when playing their games. In addition, girls in groups use different patterns of communication and different styles of communication when playing together. Tannen and her colleagues have found that young girls often play in small groups or in pairs. They play less often in large groups or teams outside the home. Girls' play is not so hierarchically ordered as boys' play is. In their most frequent games, like hopscotch and jump rope, every girl gets a chance to play. In many of their play activities, such as playing house, there are no "winners"

or “losers.” Researchers also found that girls usually don’t give many direct orders or commands to their playmates; they express their preferences as suggestions. Girls often say to their playmates, “Let’s do this ... or that.” Boys, on the other hand, are more direct in ordering their playmates to do this or that. Tannen is quick to point out that North American boys as well as girls want to get their own way, and want other children to do what they want them to do; however, the boys and girls try to get their playmates to do what they want them to do in different ways. Another researcher, compared boys and girls engaged in two task-oriented activities. The boys were making slingshots in preparation for a fight. The girls were making jewelry; they were making rings for their fingers. She noted that the boys’ activity group was hierarchically arranged. The “leader” told the other boys what to do and how to do it. The girls making the jewellery were more egalitarian. Everyone made suggestions about how to make the jewellery, and the girls tended to listen and accept the suggestions of the other girls in the group.

Researchers are not suggesting that girls never engage in some of the communication and management behaviors boys engage in. In fact, in another study, she found that when girls play house, the girl who plays the mother gives orders to the girls who play the children. But, girls ~~seem~~ to give orders to their peers less often than boys do when they play. The girls are practising parent-child relationships in the game of playing house. It’s very likely that when little boys play their games, they are also practicing the masculine roles they’re expected to assume when they grow up.

As a result of our cultural upbringing, we learn norms of behavior and patterns of communication that are often gender-based, and sometimes gender-biased. We also develop stereotypes about how and how much males and females should communicate. However, researchers have shown that many of these stereotypes actually turn out to be quite wrong.

One of the most common stereotypes that many people hold is the idea that women talk a lot, perhaps too much, and that they are always interrupting or trying to get “center stage” when someone else is talking. Actually, recent research on the influence of gender on communication has shown the exact opposite to be true in many instances.

Researchers have found that men usually produce more talk and are more likely to interrupt another speaker than women do. This particularly takes place in public settings, such as business meetings. So although women *are believed* to talk more than men, study after study

has shown that it is men who talk more. And this finding holds even for communicative interactions between very educated and successful professional men and women, such as professors, for example. Deborah Tannen, in her book *You Just Don't Understand*, cites a study of university faculty meetings. It was found that, with one exception, men professors spoke more often and, without exception, for a longer period of time than the women professors did. The men took center stage and talked from 10.66 seconds to 17.07 seconds, while the women talked from 3 to 10 seconds, on the average. Tannen points out that the women's longest speeches were still shorter than the men's shortest speeches.

Perhaps it is our social concept of what is feminine and what is masculine that reinforces the stereotype that women talk more than men, and even causes these different patterns of communication. Maybe a woman is labeled talkative or is criticized for interrupting if she does these things at all, because our culture teaches that women should be quiet if they want to be "feminine." Perhaps masculine culture encourages boys and men to dominate talk and to interrupt more often, and males who talk a lot and interrupt often are not criticized for doing so. These differences in the patterns of communication and styles of communicating are studied by researchers who study the effects of gender on communication. They study these effects in order to understand why misunderstandings occur between men and women in conversation. Often, it's because their styles and patterns of conversation are so different. It is important that we learn to recognize these differences so that we can learn to communicate better with people of the other gender.

Gender and Communication

NT

1. Define gender.

it is the social identity men & women learn as they grow up in a culture

2. What does communication between men and women depend on?

the situation the speakers find themselves in & the reasons they are communicating with other people

3. Children learn to communicate from their parents and their peers

4. How do boys achieve a higher status in a group?

by giving orders & making the other boys play by the rules & by taking center stage

5. Compare and contrast the communication patterns in boys' games and girls' games.

boys play outside in large hierarchical groups / girls play at home in small groups or pairs (not large groups) / boys have winners & losers / No winners or losers in girls' games / there is a leader giving orders in boys' games / girls make suggestions

6. How does the structure of the girls' game 'house' differ from the structure of other girls' games like 'hopscotch' and 'jump rope'?

the girl who plays the mother gives orders to the girls who play the children

7. According to Tannen, men's shortest speeches were longer than women's longest speeches in public settings.

8. Explain how culture may label men and women in terms of communication.

If women want to be feminine, they are expected to be quiet & are labeled talkative or criticised for interrupting if they do, whereas men are encouraged to dominate talk or interrupt more often. This is due to their masculine culture.

9. Why do researchers study the effects of gender on communication?

to understand why misunderstandings occur between men & women in conversation

MEDICINE

The Brain Killer

A still baffling ailment called Alzheimer's is the fourth leading cause of death for adults in the U.S. and may afflict 14 million by the year 2050

Toward the end, every move had to be plotted out for her, as if life had become a script she couldn't learn. "We are going to the dining-room table, we are going to have lunch," the nurses would say. And when she had to negotiate the tiny step to the bathroom: "Now we are coming to the step, lift up your left foot." Sometimes she missed it on the first try. "But quite often," her daughter reported, "she'll be able to do it."

It was not a towering achievement for the woman who had once effortlessly swirled across movie palace floors, matching Fred Astaire step for dazzling step. But for Rita Hayworth at 65, it in the mists of the degenerative brain disorder known as Alzheimer's disease, it repre-

sented a last, faint flicker of awareness. Three years later, bedridden and speechless, she died.

Once it sets in, Alzheimer's offers no reprieve. Whether it afflicts a forgotten movie goddess or the neighbor who quietly faded behind the upstairs curtains years ago, the disease proceeds relentlessly, stripping victims of their humanity before it takes their lives. As it destroys brain cells, first memory goes, then cognition, then physical functioning. Finally, only a shell of the person is left, evoking every child's nightmare of a parent's decline to incompetence.

There are relatively few day-care facilities for Alzheimer's sufferers

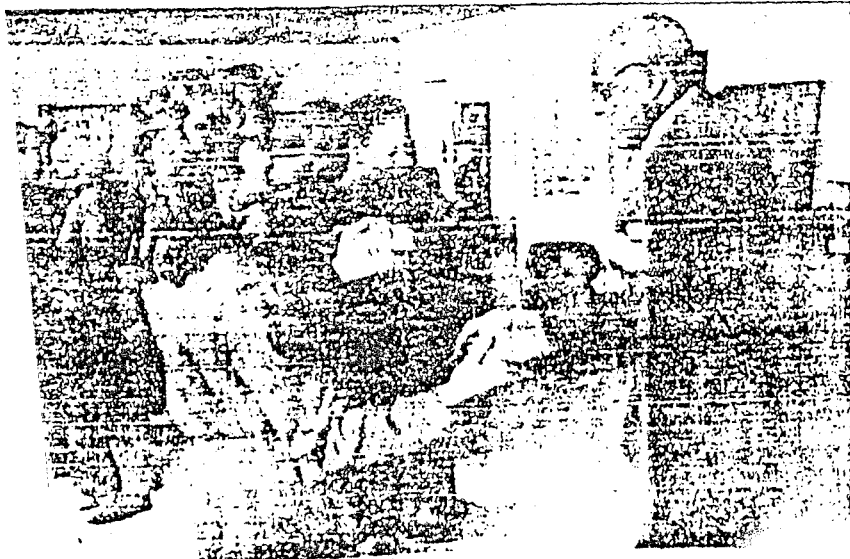
"The man I know today is not the man I knew as my father, though I still love him," says Gail Deaton, who drives 75 miles from Oklahoma City to her parents' home every week-

end to help her mother take care of him. "He is in the last phases of the illness," Deaton says. "We have to feed him, dress him, toilet him." Families of Alzheimer's victims face such tasks every day, and the impact can be financially as well as emotionally devastating. Not without reason, they are called the "hidden victims" of the disease (following story).

Until recently, about 2.5 million Americans were believed to be suffering from Alzheimer's. Then, last month, came word that the dimensions of the problem are more overwhelming than anyone had realized. Researchers at Boston's Brigham and Women's Hospital announced the results of a major study that found more than 10 percent of Americans 65 and older—and nearly half of those over 85—were suffering from "probable" Alzheimer's. According to the National Institute on Aging (NIA), that means as many as 4 million victims nationwide, almost doubling previous estimates. Moreover, with the steady growth of the over-65 population and the expected quadrupling of the over-85 group, the NIA projects a alarming 14 million Alzheimer's victims by the year 2050. The numbers alone make it "one of the biggest public-health dilemmas we've ever encountered," says NIA deputy director Gene Cohen.

Beyond sheer numbers, the revised estimates carry deeply troubling implications

JACQUES CHENET—NEWSWEEK





Number 51

JONI BURTON-WHEELER

for a health-care system already strained to the limit. According to an NIA analysis, the cost in 1985 alone of Alzheimer's and other dementias, such as those caused by strokes, was \$88 billion. The figure takes in, among other items, nursing-home care and social services, as well as some costs of assisting families at home. In Cohen's view, the prospect of a growing Alzheimer's "epidemic" in the coming decades calls for the equivalent of a Manhattan Project to tackle the disease.

"With an effort of sufficient magnitude," he says, "we could unlock the mystery behind it by the end of the century." The reality, however, is that this year the National Institutes of Health managed to allocate only \$123.4 million for Alzheimer's research—at that, more than triple the amount spent five years ago.

Alzheimer's disease is known to strike mainly the aging. But no one really knows why it strikes. While researchers have a gallery of suspects from genetic links to environmental toxins, there is as yet no known cause. It cannot even be diagnosed with absolute certainty

A health aide feeds Minna Pollock in her home, easing the family burden

except in autopsies, which pinpoint the distorted brain cells that mark the disease. And although some of its symptoms can be treated, there is not an inkling of a cure.

More is known about what the disease isn't: it is not, as was long believed, a result of the normal aging process, simple "senility," or hardening of the arteries, but rather, a pathological condition that causes lesions of the brain. Named after Alois Alzheimer, the German neurologist who first described its charac-

teristic "neurofibrillary tangles" in 1904, the disease was originally thought to be rare, and remained relatively unknown until recently as the 1970s. (A new biography of Rita Hayworth says her mental deterioration was attributed at first to alcoholism. But in the past decade it emerged as the fourth leading killer of adults, taking more than 100,000 lives annually. As a result, even a momentary memory lapse now triggers fears of Alzheimer's, though the real warning signals are less fleeting. Caregivers are terrified that the disease runs in families. "Is it hereditary, am I going to get it?" is often an issue," says Lissa Kapust, a social worker at Boston's Beth Israel Hospital who works mostly with relatives of Alzheimer's patients. "Family members are looking not only at the patient but also at what may be their own future." Actually, 10 to 30 percent of Alzheimer's cases are believed to be hereditary; the rest are of the so-called sporadic variety.

Families are bearing most of the burden of caring for patients. Roughly 70 percent of Alzheimer's victims remain at home, and eventually they need round-the-clock attention. Even in earlier stages they are prone to wandering, and outbursts of anger or violence. Later they become incontinent, and cannot feed or dress themselves. When death finally comes, it is usually from infections or pneumonia, a result of being bedridden for long periods.

Endless funeral: There is strong new evidence for the notion that each case of Alzheimer's claims at least two victims. Studies show that the impact of prolonged emotional and physical stress affects the immune system of caregivers, making them more vulnerable to infectious disease. They are often of advanced age to begin with, and may have chronic problems of their own, such as arthritis. Meanwhile, patients survive, progressively worsening, for an average of 10 years—an ordeal that has been compared to watching an endless funeral. In a way, the victim ceases to exist

but continues to live. "You go through episodes of wanting to be relieved of it, the horror that goes on day after day, night after night," says Jack Pollock, a 56-year-old Brooklyn, N.Y., high-school principal whose wife was diagnosed with Alzheimer's 10 years ago and is unable to speak or recognize anyone. "You often feel a desire that the person die—and then you feel like a monster for entertaining such thoughts."

Pollock, even so, is among the luckier caregivers. He has been able to employ two health aides to tend to his wife during the day, allowing him to keep his job—and his sanity. For that he

Facing the Facts

Approximately 10 percent of people 65 and over have "probable" Alzheimer's; 47 percent of those over 85 have the disease.

\$123.4 million in federal funds was allocated to Alzheimer's research in 1989, compared with \$5.1 million allocated in 1978.

An estimated 10 to 30 percent of Alzheimer's patients have the type that is inherited.

Stress from looking after Alzheimer's patients makes caregivers more vulnerable to infectious disease.

SOURCES: NATIONAL INSTITUTE ON AGING, ALZHEIMER'S ASSOCIATION

pays \$350 a week, plus social security, unemployment insurance and the added expense of diapers and medication. Edward Truschke, president of the Chicago-based Alzheimer's Association, estimates the overall cost of caring for a patient at home ranges from \$18,000 to \$20,000 a year. Nursing-home care may cost as much as \$36,000 in some states, but on average about \$25,000, he says. Medicare doesn't cover such chronic-care costs, and Medicaid imposes a means test. To qualify, families may have to spend themselves into poverty.

Exhausting vigil: Clearly, there is a desperate need for day-care and respite-care facilities to provide relief from the exhausting vigil. But despite some well-meaning efforts, there are relatively few facilities in place. Nationally, around 2,000 day-care facilities under various auspices are open to the elderly, and most will take Alzheimer's victims only in the very early stages of the disease. The Robert Wood Johnson Foundation is funding 19 special Alzheimer's day-care programs around the country. Generally they provide exercise programs or practical help, such as toilet training, but have no medical staff.

For "difficult" patients who need to be constantly watched or restrained, there is almost nothing out there. Neither nursing homes nor day-care facilities will take Alzheimer's patients who are disruptive or need to be restrained in any way. And day care is not free, in any case. A Johnson-funded program in Syracuse, N.Y., for example, charges \$27 a day including meals (not quite covering its own costs), and accepts neither Medicare nor Medicaid.

That is not on the whole a promising picture for an illness that cuts down one out of every 10 Americans over 65 and threatens to swamp the country's health-care capacities in the next century. Fears about Alzheimer's have been one driving force behind the current long-term-care debate in Washington. The potentially disastrous costs of the disease give budget-conscious lawmakers pause. They worry about the so-called "woodwork effect"—the unknown numbers silently providing care for their relatives who would gladly turn to professional help if the government, or private health insurers, were to pick up the tab.

Yet if there is an argument over where a family's responsibility for bill-paying ends and the government's begins, there seems no question about the government's obligation to support research. Cohen's call for a "Manhattan Project," to galvanize the research effort may seem hyperbolic, but it is not inappropriate. There is a bomb ticking away in the Alzheimer's dilemma, he believes. "As we see the staggering numbers ahead, the ticking of the clock gets louder." It may be ticking, in fact, for all of us.

DAVID GELMAN with MARY HAGER
in Washington and VICKI QUADE in Chicago

The Brain Killer

Scanning (20 minutes)

1. Since she could never recall what to do next, every move had to be _____
_____.
2. Rita Hayworth died of _____,
namely _____ at the age of _____.
3. Can Alzheimer's disease be treated? _____.
4. When a patient is afflicted by Alzheimer's he loses _____
first which is followed by _____ and finally death is
preceded by loss of _____ when only "the shell"
of the person is left.
5. Who are "the hidden victims" of the disease? Why are they called
this? _____
_____.
6. While _____ were thought to be suffering from Alzheimer's,
this figure has been found to be _____. The figure
is expected to become _____ by the year 2050 because
of _____
_____.
7. Does Cohen think the cure for the disease may be found by the year
2000? _____.
8. How can Alzheimer's be pinpointed conclusively? _____.
9. Alzheimer's is _____ that leads to _____
_____.
10. How rare is Alzheimer's? How do you know? _____
_____.

APPENDIX E

A Brief History of Language Teaching

It has been estimated that some sixty percent of today's world population is multilingual. Both from a contemporary and a historical perspective, bilingualism or multilingualism is the norm rather than the exception. It is fair, then, to say that throughout history foreign language learning has always been an important practical concern. Whereas today English is the world's most widely studied foreign language, five hundred years ago it was Latin, for it was the dominant language of education, commerce, religion, and government in the western world. In the sixteenth century, however, French, Italian, and English gained in importance as a result of political changes in Europe, and Latin gradually became displaced as a language of spoken and written communication.

As the status of Latin diminished from that of a living language to that of an "occasional" subject in the school curriculum, the study of Latin took on a different function. The study of classical Latin (the Latin in which the classical works of Virgil, Ovid, and Cicero were written) and an analysis of its grammar and rhetoric became the model for foreign language study from the seventeenth to the nineteenth centuries. Children entering "grammar school" in the sixteenth, seventeenth, and eighteenth centuries in England were initially given a rigorous introduction to Latin grammar, which was taught through note learning of grammar rules, study of declensions and conjugations, translation, and practice in writing sample sentences, sometimes with the use of parallel bilingual texts and dialogue (Kelly 1969; Howatt). Once basic proficiency was established, students were introduced to the advanced study of grammar and rhetoric. School learning must have been a deadening experience for children, for lapses in knowledge were often met with brutal punishment. There were occasional attempts to promote alternative approaches to education; Roger Ascham and Montaigne in the sixteenth century and Comenius and John Locke in the seventeenth century, for example, had made specific proposals for curriculum reform and for changes in the way Latin taught (Kelly 1969; Howatt)

1984) , but since Latin (and, to a lesser extent, Greek) had for so long been regarded as the classical and therefore most ideal form of language, it was not surprising that ideas about the role of language study in the curriculum reflected the long-established status of Latin.

As “modern” languages began to enter the curriculum of European schools in the eighteenth century, they were taught using the same basic procedures that were used for teaching Latin. Textbooks consisted of statements of abstract grammar rules, lists of vocabulary, and sentences for translation. Speaking the foreign language was not the goal, and oral practice was limited to students reading aloud sentences they had translated. These sentences were constructed to illustrate the grammatical system of the language and consequently bore no relation to the language of real communication. Students laboured over translating sentences like the following:

The philosopher pulled the lower jaw of the hen.

My sons have bought the mirrors of the Duke.

The cat of my aunt is more treacherous than the dog of your uncle

By the nineteenth century, this approach based on the study of Latin had become the standard way of studying foreign languages in schools. A typical textbook in the mid-nineteenth century thus consisted of chapters or lessons organised around grammar points. Each grammar point was listed, rules on its use were explained, and it was illustrated by sample sentences. This approach to foreign language teaching became known as the Grammar-Translation Method.

The Grammar-Translation Method

As the names of its leading exponent suggest (Johann Seidenstücker, Karl Plötz, H.S. Ollendorf, and Johann Meidinger), Grammar-translation was the offspring of German scholarship, the object of which, according to one of its less charitable critics, was “to know everything about something rather than the thing itself” (W.H.D. Rouse, quoted in Kelly 1969: 53). Grammar-Translation was in fact the first known in the United States as the Prussian Method.

Grammar-Translation dominated European and foreign language teaching from the 1840s to the 1940s, and in modified form it continues to be widely used in some parts of the world today. At its best, as Howatt (1984) points out, it was not necessarily the horror that its critics depicted it as. Its worst excesses were introduced by those who wanted to demonstrate that the study of French or German was no less rigorous than the study of classical languages. This resulted in the type of Grammar-Translation courses remembered with distaste by thousands of school learners, for whom foreign language learning meant a tedious experience of memorising endless lists of unusable grammar rules and vocabulary and attempting to produce perfect translations of stilted or literary prose. Although the Grammar-Translation Method often creates frustration for students, it makes few demands on teachers. It is still used in situations where understanding literary texts is the primary focus of foreign language study and there is little need for a speaking knowledge of the language. Contemporary texts for the teaching of foreign languages at college level often reflect Grammar-Translation principles. These texts are frequently the products of people trained in literature rather than in language teaching or applied linguistics. Consequently, though it may be true to say that the Grammar-Translation Method is still widely practiced, it has no advocates. It is a method for which there is no theory. There is no literature that offers a rationale or justification for it or that attempts to issues in linguistics, psychology, or educational theory.

In the mid-and late nineteenth century opposition to the Grammar-Translation Method gradually developed in several European countries. This Reform Movement, as it was referred to, laid the foundations for the development of new ways of teaching languages and raised controversies that have continued to the present day.

Language teaching innovations in the nineteenth century

Toward the mid-nineteenth century several factors contributed to a questioning and rejection of the Grammar-Translation Method. Increased opportunities for communication among Europeans created a demand for oral proficiency in foreign languages. Initially this created a market for conversation books and phrase books

intended for private study, but language teaching specialists also turned their attention to the way modern languages were being taught in secondary schools. Increasingly the public education system was seen to be failing in its responsibilities. In Germany, England, France, and other parts of Europe, new approaches to language teaching specialists, each with a specific method for reforming the teaching of modern languages. Some of these specialists, like C. Marcel, T. Prendergast, and F. Gouin, did not manage to achieve any lasting impact, though their ideas are of historical interest.

The Frenchman C. Marcel (1873-1896) referred to child language learning as a model for language teaching, emphasised the importance of meaning in learning, proposed that reading be taught before other skills, and tried to locate language teaching within a broader educational framework. The Englishman T. Prendergast (1806-1886) was one of the first to record the observation that children use contextual and situational cues to interpret utterances and that they use memorised phrases and "routines" in speaking. He proposed the first "structural syllabus," advocating that learners be taught the most basic structural patterns occurring in the language. In this way he was anticipating an issue that was to be taken up in the 1920s and 1930s.

The Frenchman F. Gouin (1831-1896) is perhaps the best known of these mid-nineteenth century reformers. Gouin developed an approach to teaching a foreign language based on his observations of children's use of language. He believed that language learning was facilitated through using language to accomplish events consisting of a sequence of related actions. His method used situations and themes as ways of organising and presenting oral language-the famous Gouin "series", which includes sequences of sentences related to such activities as chopping wood and opening the door. Gouin established schools to teach according to his method, and it was quite popular for a time. Gouin's emphasis on the need to present new teaching items in a context that makes their meaning clear, and the use of gestures and actions to convey the meanings of utterances, are practices that later became part of such approaches and methods as Situational Language Teaching and Total Physical Response.

The work of individual language specialists like these reflects the changing climate of the times in which they worked. Educators recognised the need for

speaking proficiency rather than reading comprehension, grammar or literary appreciation as the goal for foreign language programs; there was an interest in how children learn languages, which prompted attempts to develop teaching principles from observation of ;(or more typically reflections about) child language learning. But the ideas and methods of Marcel, Prendergast, Gouin, and other innovators were developed outside the context of established circles of education and hence lacked the means for wider dissemination, acceptance, and implementation. They were writing at a time when there was not sufficient organisational structure in the language teaching profession (i. e. , in the form of professional associations, journals and conferences) to enable new ideas to develop into an educational movement. This began to change toward the end of the nineteenth century, however, when a more concerted effort arose in which the interests of reform-minded language teachers, and linguists, coincided. Teachers and linguists began to write about the need for new approaches to language teaching, and through their pamphlets, books, speeches, and articles, the foundation for more widespread pedagogical reforms was laid. This effort became known as the Reform Movement in language teaching.

The Reform Movement

Language teaching specialists like Marcel, Prendergast, and Gouin had done much to promote alternative approaches to language teaching, but their ideas failed to receive widespread support or attention. From the 1880s, however, practically minded linguists like Henry Sweet in England, Wilhelm Wietor in Germany, and Paul Passy in France began to provide the intellectual leadership needed to give reformist ideas greater credibility and acceptance. The discipline of linguistics was revitalised. Phonetics -the scientific analysis and description of the sound systems of languages - was established, giving new insight into speech processes. Linguists emphasised that speech, rather than the written word, was the primary form of language. The International Phonetic Association was founded in 1886, and its International Phonetic Alphabet (IPA) was designed to enable the sounds of any language to be accurately transcribed. One of the earliest goals of the association was to improve the teaching of modern languages. It advocated

1. the study of the spoken language;
2. phonetic training in order to establish good pronunciation habits;
3. the use of conversation texts and dialogues to introduce conversational phrases and idioms;
4. an inductive approach to the teaching of grammar;
5. teaching new meanings through establishing associations within the target language rather than by establishing associations with the mother tongue.

Viotor, Sweet, and other reformers in the late nineteenth century shared many beliefs about the principles on which a new approach to teaching foreign languages should be based although they often differed considerably in the specific procedures they advocated for teaching a language.

These principles provided the theoretical foundations for a principled approach to language teaching, one based on a scientific approach to the study of language learning. They reflect the beginnings of the discipline of applied linguistics –that branch of language study concerned with the scientific study of second and foreign language teaching and learning. The writings of such scholars as Sweet, Viotor, and Passy provided suggestions on how these applied linguistics principles could best be put into practice. None of these proposals assumed the status of a method, however, in the sense of a widely recognised and uniformly implemented design for teaching a language. But parallel to the ideas put forward by members of the Reform Movements was interest in developing principles for language teaching out of naturalistic principles of language learning, such as are seen in first language acquisition. This led to what have been termed natural methods and ultimately led to the development of what came to be known as the Direct Method.

The Direct Method

Gouin had been one of the first of the nineteenth century reformers to attempt to build a methodology around observation of child language learning. Other reformers toward the end of the century likewise turned their attention to naturalistic principles of language learning, and for this reason they are sometimes referred to as advocates of a “natural” method. In fact at various times throughout the history of language

teaching, attempts have been made to make second language learning more like first language learning. In the sixteenth century, for example, Montaigne described how he was entrusted to a guardian who addressed him exclusively in Latin for the first years of his life, since Montaigne's father wanted his son to speak Latin well. Among those who tried to apply natural principles to languages classes in the nineteenth century L. Sauveur (1826-1907), who used intensive oral interaction in the target language, employing questions as a way of presenting and eliciting language. He opened a language school in Boston in the late 1860s, and his method soon became referred to as the Natural Method.

Sauveur and other believers in the Natural Method argued that a foreign language could be taught without translation or the use of the learner's native tongue if meaning was conveyed directly through demonstration and action. The German scholar F. Franke wrote on the psychological principles of direct association between forms and meanings in the target language (1884) and provided a theoretical justification for a monolingual approach to teaching. According to Franke, a language could best be taught by using it actively in the classroom. Rather than using analytical procedures that focus on explanation of grammar rules in classroom teaching, teachers must encouraged direct and spontaneous use of the foreign language in the classroom. Learners would then be able to induce rules of grammar. The teacher replaced the textbook in the early stages of learning. Speaking began with systematic attention to pronunciation. Known could be used to teach new vocabulary, using mime, demonstration and pictures.

These natural language learning principles provided the foundation for what came to be known as the Direct Method, which refers to the most widely known of the natural methods. Enthusiastic supporters of the Direct Method introduced it in France and Germany (it was officially approved in both countries at the turn of the century), and it became widely known in United States through its use by Sauveur and Maximilian Berlitz in successful commercial language schools. (Berlitz, in fact, never used the term; he referred to the method used in his schools as the Berlitz Method.) In practice it stood for the following principles and procedures:

1. Classroom instruction was conducted exclusively in the target language.

2. Only everyday vocabulary and sentences were taught.
3. Oral communication skills were built up in a carefully graded progression organised around question-and-answer exchanges between teachers and students in small, intensive classes.
4. Grammar was taught inductively.
5. New teaching points were introduced orally.
6. Concrete vocabulary was taught demonstration, objects, and pictures; abstract vocabulary was taught by association of ideas.
7. Both speech and listening comprehension were taught.
8. Correct pronunciation and grammar were emphasised.

The Direct Method method was quite successful in private language schools, such as those of the Berlitz chain, where paying clients had high motivation and the use of native-speaking teachers was the norm. But despite pressure from proponents of the method, it was difficult to implement in public secondary school education. It overemphasised and distorted the similarities between naturalistic first language learning and classroom foreign language learning and failed to consider the practical realities of the classroom. In addition, it lacked a rigorous basis in applied linguistic theory, and for this reason it was often criticised by the more academically based proponents of the Reform Movement. The Direct Method represented the product of enlightened amateurism. It was perceived to have several drawbacks. First, it required teachers who were native speakers or who have nativelike fluency in the foreign language. It was largely dependent on the teacher's skill, rather than on a textbook, and not all teachers were proficient enough in the foreign language to adhere to the principles of the method. Critics pointed out that strict adherence to Direct Method principles was often counterproductive, since teachers were required to go to great lengths to avoid using the native tongue, when sometimes a simple brief explanation in the student's native tongue would have been a more efficient route to comprehension.

The Harvard psychologist Roger Brown has documented similar problems with strict Direct Method techniques. He described his frustration in observing a teacher performing verbal gymnastics in an attempt to convey the meaning of Japanese

words, when translation would have been a much more efficient technique to use (Brown 1973:5).

Although the Direct Method enjoyed the popularity in Europe, not everyone had embraced it enthusiastically. The British applied linguist Henry Sweet had recognised its limitations. It offered innovations at the level of teaching procedures but lacked a through methodological basis. Its main focus was on the exclusive use of the target language in the classroom, but it failed to address many issues that Sweet thought more basic. Sweet and other applied linguists argued for the development of sound methodological principles that could serve as the basis for teaching techniques. In the 1920s and 1930s applied linguists systemised the principles proposed earlier by the Reform Movement and so laid the foundations for what developed into the British approach to teaching English as a foreign language. Subsequent developments led to Audiolingualism in the United States and the Oral Approach or Situational Language Teaching in Britain.

What became of the concept of method as foreign language teaching emerged as a significant educational issue in the nineteenth and twentieth centuries? We have seen from this historical survey some of the questions that prompted innovations and new directions in language teaching in the past:

1. What should the goals of language teaching be? Should a language course try to teach conversational proficiency, reading, translation, or some other skill?
2. What is the basic nature of language, and how will this affect teaching method?
3. What are the principles for the selection of language content in language teaching?
4. What principles of organisation, sequencing, and presentation best facilitate learning?
5. What should the role of the native language be?
6. What process do learners use in mastering a language, and can these be incorporated into a method?
7. What teaching techniques and activities work best and under what circumstances?

A Brief History of Language Teaching

Answer the following questions:

1. In the 16th, 17th, and 18th centuries in England, students were taught Latin grammar at the preliminary level through *note learning of grammar rules* *study of declensions and conjugations* *translation* and *practice in writing sample sentences, sometimes with the use of bilingual texts and dialogues*

2. The students who had to learn languages through the Grammar Translation method were frustrated because they had to:

a)

b)

3. The need for oral proficiency in foreign languages was a result of

.....and led first to

then it also created an interest in

4. The ideas and methods of Marcel, Prendergast and Gouin did not develop into an educational movement because at their time, the language teaching profession did not have

..... such as

..... and

5. After 1880, linguists who believed that was more

important than, helped to establish

phonetics, which is

6. According to Franke, learners will be able to if they are encouraged to use the language and in the classroom.

7. The Direct Method, widely used in the U.S. in schools, advocates the teaching of vocabulary, a(n) approach to grammar and putting stress on

8. The Direct Method was not very successful in public secondary school education because

- a).....
- b).....
- c).....

9. The principles of the Reform Movement were by the applied linguists and this led to an approach known as and later as in the U.S. and in Britain.

Vocabulary

Find a word in the text which means the same as the followings:

1. point of view (p.1)
2. changed or altered (p.5, 11. 1-12)
3. strong effect, influence (p.7)
4. encourage, support (p.8, 11. 1-12)
5. give rise to, inspire (p.9, 11. 1-12)
6. put into practice, carry out (p.12)
7. express, communicate (p.14)

Referrals

1. that (par. 2 L.1)
2. they (par. 3 L.2)
3. its (par. 4 L.1)
4. it (par. 5, L.20)
5. its (par. 7, L.8)
6. their (par. 8, L.21)
7. they (par. 11, L.3)
8. it (par. 15, L.9)
9. its (par. 18, L.4)

The Argumentative Essay

Write an argumentative essay defending one of the two points of view given below on drug use.

Soft drugs like marijuana should be legalized

1. Drugs are a 'forbidden fruit' making it attractive to the young. Legalization would remove this temptation.
2. Crime rate is high because drugs are expensive.
Addicts turn to crime to finance the habit, half the street crimes are drug related – 'American Bar Association' The crime rate would be reduced if drugs are made legal.
3. Legalization would remove the pusher - the pusher (a seller of drugs) often the initiator of drugs to the young, possibly has criminal tendencies, tends to use blackmail for unpaid drug money.
4. Drugs big money – inevitable that poorly paid police and government officials become involved eg. Susurluk incident. Legislation would remove the big money syndrome.
5. Medical evidence has shown marijuana to be a palliative to those terminally ill eg. cancer.
6. Evidence suggests that cigarettes and alcohol are more harmful than marijuana.

Soft drugs like marijuana should not be legalized

1. Legalizing drugs increases users and addiction
2. Most addicts given free drugs by pushers – addict becomes captive customer.
3. A need for money to satisfy drug craving leads to crime (drugs expensive)
4. Five joints of marijuana are equal to sixteen cigarettes – leading to health problems such as those caused by smoking.
5. A danger in moving from soft drugs to hard drugs eg. heroin.
6. Using shared needles a danger to health – a high incidence of being HIV positive resulting.
7. Drugs lead to escaping from life instead of facing life's realities.

Argumentative Essay

Advertising

Below are two introductory paragraphs on advertising. One is arguing for advertising and the other against. Read them to see how they put forward their different points of view.

Against

Advertising has always been a source of argument. Ever since men began selling to each other it has been accused of exaggeration and deception. It, also, encourages people to buy goods they do not need.

For

We all enjoy some advertisements although they may not necessarily influence our buying habits. It would be impossible for society to survive if manufacturers did not sell their goods. Advertising helps them to do this.

Look at the following points 'for' or 'against' advertising. Choose one of them and write an argumentative essay 'for' or 'against' advertising.

Advertising is a good thing

- a. Gives the shopper useful information, the shopper knows what he is buying eg a computer, cars toothpaste etc.
- b. Creates competition – manufacturers make their products to a high standard, competition reduces prices.
- c. Advertisements pay for I TV programmes (no need to buy a licence as in England) keeps down the cost of newspapers and magazines.
- d. Encourages one to buy attractive things, giving pleasure to life and contributes to a country's economic growth.
- e. Advertising is an essential feature of a free trade society.

Advertising is a bad thing

- a. Misleads shoppers by making false claims – what are the new and exciting formulas in shampoos or detergents.
- b. Sets out to deceive – do we ever see a motor car in a traffic - jam .
- c. Precious real information given – what about preservatives, additives and colouring in food, or aid's to slimming, drugs and patent medicines.
- d. Use of family life to sell – professional, successful middle class, two children a boy and a girl (the boy always the elder)
- e. Banks always portrayed as nice friendly places, willing to lend money easily – no mention of the consequences is made should the borrower default on repayments.
- f. Spoils I TV programmes by interrupting films etc. leading to a loss of interest, and concentration.
- g. Advertising is anti-democratic as it makes it harder for the shopper to choose freely.

Appendix G

Three Sample Compositions

A Paper Rated as "Not Adequate"

Nowadays a lot of countries tend to develop their tourism's incomes, and therefore tourism called the factory without chimney. Turkey, which undoubtedly needs foreign money, tries to increase the number of foreign tourists coming to Turkey. What are likely to do in order to increase this number.

At first, much more and better advertising should do in foreign countries and the information offices should open to inform the people to decide to come Turkey. Secondly, improve facilities, which are hotels, transportation and communication. Increase the number of hotels, similarly the number of public transportation which, improve the lines of communication. Thirdly which is important as two others is training of personnel. This is also a basic need of tourism, because the tourist will want to see in front of him a skilled guides or a skilled hotel managers. The new school will open in order to train skilled personnel and as well as theoretic knowledges, practice must be given them.

The countries which are made available these three basic need for tourists have already improved their tourism's incomes. Spain is a case in point of Greece. Although Turkey needs this income; it didn't do any real attempts to achieve it. In fact all of them should have already been done, till today. However it is late, it can be begin without losing any time.

A Paper Rated as "Adequate"

Tourism is now becoming a major industry throughout the world. For many countries their tourist trade is an essential source of their revenue.

All countries have their own particular attractions for tourists and this must be kept in mind when advertising Turkey abroad. For example Turkey, which wants to increase the number of foreign tourists coming must advertise its culture and sunshine.

Improving facilities like hotels, transportation and communication play an important role on this matter. More hotels can be built and available ones can be kept clean and tidy. New and modern transportation systems must be given to foreign tourists and one more, the communication system must work regularly to please these people.

Tourists don't want to be led around like sheep. They want to explore for themselves and avoid the places which are packed out with many other tourists. Because of that there must be their trained guides on their tours through anywhere and on the other hand hotel managers must be well trained. They must keep being kind to foreign tourists and must know English as well.

If we make tourists feel comfortable in these facts, tourism will increase and we will benefit from it.

Paper Rated as "More Than Adequate"

A nation can't make improvements, if it doesn't let the minds of their people breathe and expand to understand more about life than what is at the end of the street, this improvement can be made by means of tourism.

There are several ways to attract more people to our country. First of all, advertisements and information take an important place. These advertisements and information should be based on the qualities of that place without exaggeration. The more time passes and the more information tourists gather about the country, the more assured they can be that it will be a good experience. People travel one place to another in order to spend their holiday, to see different cultures or to attend conferences. All of these necessitate facilities. It is important to make some points clear. Hotel, transportation and communication facilities are a case in point. To some extent, we can minimize the difficulties by means of money. Furthermore, this situation does not only depend on the financial situation, but also behaviors towards tourists. Especially, a developing country should keep in mind the challenge of the future rather than the mistakes of the past, in order to achieve this, the ways of training of personnel may be found. The most important problem faced by many of countries is whether the decisions that must be made are within the capabilities of their education system. Educating guides and hotel managers are becoming more and more important.

As a result, it should once more be said that, we may increase the number of foreign tourists coming to Turkey by taking some measures. Advertisement, information, improving facilities and training personnel may be effective, but also all people should be encouraged to contribute this event.

Appendix H

STUDENTS' QUESTIONNAIRE

Part I

1. Please write your students' number:.....
2. Department: FLED ☐ LL ☐ TRANS ☐
3. Gender: Male ☐ Female ☐
4. Age:
5. How many questions did you correctly answer on the ÖYS English Test?
6. Which school did you graduate from?

Part II

1. Please, order the parts of BUEPT according to their level of difficulty (1=the easiest; 3=most difficult)

LISTENING ☐ READING ☐ WRITING ☐

2. In your opinion, what was the main goal of prep classes? (Note: You may tick more than one choice).

☐ To teach students how to pass BUEPT

☐ To prepare students for their future academic studies

☐ To develop students' ☐ reading skills ☐ listening skills ☐ writing skills

☐ speaking skills

☐ To improve the level of students' English proficiency

☐ Other (please explain)

.....

.....

3. In your opinion, should there be a test measuring your level of academic English proficiency before you begin your freshman year studies?

Yes ☐ No ☐ I don't know ☐

Why do you think so?.....
.....
.....
.....

4. In your opinion, did BUEPT accurately reflect your level of academic English proficiency?

Yes ☐ No ☐ I don't know ☐

Why do you think so?
.....
.....
.....

5. In your opinion, did BUEPT properly predict your freshman academic success?

Yes ☐ No ☐ I don't know ☐

Why do you think so?.....
.....
.....
.....