



T. C. ANADOLU ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

THE COMMENTS OF EFL TEACHERS ON PREPARING ACHIEVEMENT TESTS AND AN ANALYSIS OF A SAMPLE TEST

(A Study at the Engineering Faculty of)
Anadolu University

(Yüksek Lisans Tezi)

Ateş DALYAN

Eskişehir, 1990

Anadolu Üniversitesi
Sosyal Bilimler Enstitüsü

ACKNOWLEDGEMENTS

I would like to express my gratitude and heartfelt thanks to my thesis supervisor Dr. Ahmet KONROT under whose guidance I worked and whose ideas, suggestions, comments and criticisms were invaluable.

I wish to express my special thanks to Dr. Gül DURMUŞOĞLU and Dr. Zülal BALPINAR who had given their time for contributions and great efforts to the various linguistic studies of Master of Arts Programme in English Language Teaching.

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

Special thanks must also go to my old friends and colleagues Semih-Melek Şahinel and Hakan-Mahinur Arslan who encouraged me in my studies on this thesis and Atilâ Onur, who typed this thesis.

Finally, I would like to express my thanks to my parents for their understanding, patience and moral support during the course of this study.

ABSTRACT

Recent years have witnessed a number of important developments in English language testing. Consequently, the potential influence of testing on teaching has assumed a new significance. As a natural result of this, there is a growing need for more effective testing to establish appropriate standards of English and to measure accurately student performance with respect to them in every language instruction program. That's why, this thesis focuses upon the assessment of teachers' opinions on the present English language testing system at the Faculty of Engineering, Anadolu University. The main assumption of this study is that the findings will help to establish some suggestions for more feasible ways of increasing facility in language testing at the Faculty of Engineering. The first chapter introduces background to the problem by pointing out the importance of language testing issue together with the purpose of language testing, some definitions of language tests, the problems that exist and the aim of the study. Chapter II reviews literature relevant to the study. In Chapter III, the research method is explained. Chapter IV includes the results of analysis. In this

chapter, the findings determined that the subjects do not agree with each other on the appropriateness of English language tests administered at the Faculty of Engineering. In the last chapter, Chapter V, the discussion over the research results and some suggestions for further studies are presented.

ÖZET

Son yıllarda İngilizce dil testlerinde bazı önemli gelişmeler görülmektedir. Bunun sonucunda da testlerin öğretime olan yoğun etkisi yeni bir önem kazanmıştır. Bunun doğal sonucu olarak da her yabancı dil öğretim programında uygun İngilizce standartları saptayabilmek ve öğrencilerin gereksinimleri çerçevesinde başarıyı tam olarak ölçebilmek için daha etkin testlere ihtiyaç duyulmaktadır. Bu nedenle bu çalışma Anadolu Üniversitesi Mühendislik-Mimarlık Fakültesinde uygulanmakta olan İngilizce test sistemi ile ilgili olarak aynı fakültenin İngilizce öğretim elemanlarının görüşlerinin değerlendirilmesini kapsamaktadır. Bu çalışmayla elde edilen bulguların Mühendislik-Mimarlık Fakültesinde uygulanmakta olan testlerin geliştirilmesine katkıda bulunacak bazı öneriler sağlayacağı varsayılmaktadır. Çalışmanın birinci bölümünde yabancı dil testlerinin önemine değinilerek bazı test tanımları, mevcut problemler ve çalışmanın amacı verilmiştir. İkinci bölümde ise çalışmayla ilgili kaynak taraması yer alır. Üçüncü bölümde araştırma yöntemi belirtilmiştir. Dördüncü bölümde ise bulgular yer almaktadır. Bu bölümde elde edilen bulgular, Mühendislik-Mimarlık Fakültesinde

uygulanmakta olan İngilizce dil testlerinin uygunluđu konusunda İngilizce öğretim elemanlarının farklı görüşlere sahip olduklarını göstermiştir. Son bölüm olan beşinci bölümde ise bulguların yorumu ile ileriye yönelik araştırmalar için bazı öneriler yer almaktadır.

LIST OF TABLES

Table		Page
1	Teaching Experiences of the Teachers.....	72
2	Attendance of Teachers at Language Testing Courses.....	72
3	Attendance of Teachers at In-Service Courses or Seminars Concerning Language Testing.....	73
4	Opinions of the Teachers About the Present Tests.....	74
5	Opinions of the Teachers About the Appropriateness of Some Language Test Types.....	76
6	Opinions of the Teachers About the Certain Types of Tests and Their Administering Frequency.....	77
7	Opinions of the Teachers About the Purposes of Language Tests.....	79

8	Opinions of the Teachers About the Extent of Cooperation Among Teachers.....	80
9	The Extent of Cooperation Between the Teachers in Constructing the Tests.....	81
10	Opinions of the Teachers About the Calculation of the Mean of the Test Scores.....	82
11	Opinions of the Teachers About the Interpretation of the Test Results in Constructing the Forthcoming Tests.....	82
12	Opinions of the Teachers About the Use of Test Results.....	83
13	Opinions of the Teachers About the Appropriateness of Language Tests With the Aims and Practices of Language Teaching.....	85
14	Opinions of the Teachers About the Relationship Between the Language Being Tested and The Way It Is Taught.....	85
15	Opinions of the Teachers About the Appropriateness of Their Language Tests to the Course Objectives.....	86

16	Opinions of the Teachers About Some Aspects of Language Skills that the Tests Make Demands on.....	87
17	Opinions of the Teachers About the Way in which Learner is Taken into Account on Tests.....	88
18	Opinions of the Teachers About the Appropriateness of the Tests in Terms of the Testees' Knowledge, Expectations and Skills.....	88
19	Opinions of the Teachers About Their Tests' Instructions and Tasks.....	89
20	Opinions of the Teachers About the Order of Test Questions.....	90
21	Opinions of the Teachers About Avoidance of Questions that Provide More Than One Acceptable Answer.....	91
22	Opinions of the Teachers About Pre-Testing of Language Tests.....	92
23	Opinions of the Teachers About General Test Organization.....	92

24	The Appropriateness of the Midterm Test Administered at the Faculty of Engineering in 1989-90 Academic Year, From the Point of Avoided Testing Mistakes.....	96
25	Facility Value of the Items of Midterm Test.....	97
26	The Appropriateness of the Sample Test to the Checklist of Basic Principles for Constructing Achievement Tests.....	105

CONTENTS

	Page
ACKNOWLEDGEMENTS.....	ii
ABSTRACT.....	iii
ÖZET.....	v
LIST OF TABLES.....	vii
CHAPTER I	
INTRODUCTION	
1.1. Background to the problem.....	1
1.2. Problem.....	7
1.3. Purpose of the study.....	8
1.4. Importance of the study.....	10
1.5. Limitations.....	12
1.6. Assumptions.....	13
1.7. Definitions.....	13
CHAPTER II	
REVIEW OF LITERATURE	
2.0. Chapter overview.....	19

2.1. Educational uses of language tests.....	20
2.2. Approaches in language testing.....	24
2.2.1. The traditional approach.....	24
2.2.2. The psychometric-structuralist approach.....	25
2.2.3. The integrative-sociolinguistic approach.....	26
2.3. Classification of language tests.....	29
2.3.1. Achievement or attainment tests.....	31
2.3.2. Proficiency tests.....	32
2.3.3. Aptitude tests.....	33
2.3.4. Diagnostic language tests.....	34
2.4. Characteristics of tests.....	35
2.4.1. Validity.....	35
2.4.2. Reliability.....	38
2.4.3. Practicality.....	40
2.5. Construction of language tests.....	42
2.5.1. Determining the purpose of testing..	43
2.5.2. Developing the test specifications.	44
2.5.3. Selecting the appropriate item types.	46
2.5.4. Preparing relevant test items.....	48
2.5.5. Assembling the test.....	49
2.5.6. Administering the test.....	51
2.5.7. Using the results.....	55
2.6. Problems in language testing.....	56

CHAPTER III

THE RESEARCH METHOD

3.1. Research design.....	66
3.1.1. Description of questionnaire.....	67
3.2. Subjects.....	68
3.3. Data collection.....	69
3.4. Data analysis.....	70

CHAPTER IV

ANALYSIS OF RESULTS AND DISCUSSION

4.1. Evaluation and interpretation of results...	71
4.2. Some observations on the sample test.....	94

CHAPTER V

DISCUSSION AND SUGGESTIONS

5.1. Discussion.....	111
5.2. Suggestions for further studies.....	115

REFERENCES.....	117
-----------------	-----

APPENDICES

APPENDIX A-1: The Questionnaire Used to Obtain Information About the Extent of Agreement or Disagreement on the Appropriateness of English Language Tests Among Teachers at the Faculty of Engineering.....	121
--	-----

APPENDIX B-1: A Midterm Test Group B.....	128
APPENDIX B-2: A Midterm Test Group A.....	134
APPENDIX B-3: The Results of the Midterm Test....	140

CHAPTER I

INTRODUCTION

1.1. Background to the Problem

Some form of evaluation is inevitable in teaching. It is as inevitable in classroom teaching as it is in all fields of activity where judgements need to be made however simple or complex. In Education the terms Evaluation, Measurement and Testing are widely used. The term 'Evaluation' as defined by Noll and Scannell(1972:14) refers to emphasizing the role of judgement in considering a variety of types of relevant data or information in determining progress toward a goal, the worth of a program, or how to make a decision; whereas in measurement we strive to be as objective as possible and minimize the role of judgement.

On the other hand, the term "Testing" as generally used in education can come to have a rather specific and limited connotation and, in some instances, a slightly unfavorable one. Because a tester can be regarded by some persons (whether rightly or wrongly) as a technician who is more interested in the scores and statistics of the results of tests **than** in what the results mean in relation to the boys and girls who made them. In this respect, whether the process is called testing, measurement or evaluation it is not nearly as important as whether the progress or status of the learner with respect to the desired goal is being determined. Because it is obvious that the instruments or techniques should be chosen to fit the objectives and no technique is worth using unless the results it yields can be depended upon.

According to Pilliner(1973:21) examinations and tests give rise to measures of some attribute possessed by some individual. He explains the distinction between tests and examinations in terms of:

1. Duration (an examination lasts longer than a test)
2. Hierarchy of career (a university professor 'examines' his Final Honour Students in English Literature, whereas a primary school teacher 'tests' her nine-year olds in spelling)

3. The subjectiveness or objectiveness of assessment (if the assessment is subjective, the instrument is an examination, but in the latter case it is a test.)

It can be said that a clear cut distinction has not been made between tests and examinations. Because they both serve to assess the same learning outcomes in a teaching situation. Therefore, the only way to measure the outcomes of an educational program is to test students by employing one type of test or another.

According to Oller(1979:3) the problems of language testing are a large subset of the problems of educational measurement in general. About the same issue, Finocchiaro and Sako(1983:6) indicate that except the tests in mathematical sciences, all other tests make use of concepts and principles from linguistics, psychology, sociology and anthropology.

Oller(1979:11) defines language tests as any test that challenges the language ability of an examinee. On the same parallel, language tests are generally classified into various categories depending upon their format, the skills and language items being measured, and the use to be made of their results. The literature on language testing presents that most specialists have agreed on the names "proficiency",

"diagnostic", "aptitude"(prognostic or predictive) and "achievement"(or attainment) as terms of tests most frequently given.

According to Valette(1977:6) the aim of proficiency tests is to determine whether the students' language ability corresponds to specific language requirements. For example, placement tests are considered as proficiency tests.

On the other hand, Heaton(1975:6) defines diagnostic tests as the tests that enable the teacher to identify specific areas of weakness and difficulty so that he is able to plan the most appropriate remedial programme.

Another type of test is the language aptitude test. In Finocchiaro and Sako's(1983:23) explanation language aptitude tests(sometimes termed as prognostic or predictive tests) are designed to predict before beginning language study, a subject's probable capability of acquiring the language.

Finally, there are achievement or attainment tests. According to Heaton(1975:10) the achievement or attainment tests have been designed to show mastery of a particular syllabus and they may be teacher-made.

It is obvious that language testing is a complex matter. It involves elements from linguistics, psychometrics and language teaching. Why a test is set, what is

tested and when a test is set are the questions which determine the actual testing procedure.

According to Finocchiaro and Sako(1983:12) we may test in order to:

1. determine a. whether our teaching methods and techniques are in fact producing learning and b. which aspect of these are in need of revision,
2. measure students' achievement,
3. provide for ability or interest placement or grouping,
4. offer remedial help,
5. identify difficulties,
6. improve our understanding of how, where and when bilinguals use certain aspects of the languages they have learned,
7. learn more about the language acquisition process at various age levels,
8. predict the possible success or failure in language acquisition of certain individuals,
9. diagnose,
10. provide continuity of instruction,
11. assess -for specific purposes- the level of mastery(proficiency),

12. gauge the ability of teachers,
13. set realistic standards,
14. assess the effects of experimentation,
15. promote confidence of the community.

It is certain that all tests are for a purpose. A test that is made up without a clear idea of what it is for, is not good. That's why it is important to consider testing as not something that stands apart from language teaching. And it can be claimed that some language tests can be prepared better than the others. So, if some tests are better than the others it is important to use the better ones.

In sum, teachers must believe in the usefulness of, importance of and the need for tests in the teaching process and help those involved in test preparation to find ways for increasing the power of tests. In a rational teaching situation, foreign language teachers have a responsibility to find out how much improvement their students have made. So, they should make sure that the students are being scaled precisely. At the same time it can be pointed out that one of our most pressing needs today as teachers of English as a foreign language is the know-how for writing tests that will measure our pupils' achievement in English. And also these tests must be based on principles of foreign language teaching. Therefore, the problem of this study may be stated as follows:

1.2. Problem

It is certain that preparing good tests requires a good deal of time and energy. In looking over the present language testing situation, some improvement would seem to be necessary in the system at The Faculty of Engineering, Anadolu University. Because classroom teachers are expected to prepare their own language tests with limited time and overloaded schedules. Certainly, this does not mean that such overworked teachers cannot prepare any good tests. But, logically it doesn't seem to be appropriate to expect from those teachers to orientate themselves to the procedures of constructing good language tests.

At the Faculty of Engineering, language testing has always been subjected to lively debate. And a close identification of the strengths and weaknesses of the given tests is inevitable. These tests are prepared by the teachers themselves. And there are always some discussions on some points of given tests. Since the objections are heavily on the test items, the weights given to them and on the lack of interest in administering communicative oriented tests, the teaching objectives and emphases of the course or the program are also started to take place in the discussions. And, usually at the end of these discussions the immediate solution for some of us is that tests

are given that they are necessary in order to assign the students some grade for the work they have been doing because administration demands it. So, with this administrative requirement, we, without much consideration in principles of testing, prepare our tests and keep on having discussions.

Therefore, it became inevitable for me to go to my colleagues and ask what they are thinking about the testing system in the school. Are we grading our students entirely on their learning points? What do we expect these tests to achieve? Do we all agree that our testing system is consistent with the general course curriculum? Are we satisfied with the concrete results of tests? Is the language being tested in the way it is taught? By using these techniques are we actually finding what the students know? Do we test all aspects of language learning? Do we frame our questions that there is no ambiguity? Do we word our questions that there will be no guessing?

These are some questions in this study to be asked to the teachers. The answers should shed light on some acute problems of language testing at the Faculty of Engineering.

1.3. Purpose of The study

This study is intended to demonstrate the extent of agreement and disagreement on the appropriateness

of language tests among teachers and to find out some answers to the questions stated in the Problem Section (cf. 1.2) by exhibiting the present situation of tests administered at the Faculty of Engineering, Anadolu University-Bademlik Extension Campus.

The present writer of this thesis is one of the teachers of this faculty. He does not claim that he is a testing expert, or he is an ideal teacher. Nor he wants to humiliate any teacher/s, authority or institution by exhibiting some deficiencies in testing. Our main purpose is to enlighten one area of education by making some suggestions for some feasible ways of increasing facility in foreign language testing at the Faculty of Engineering.

To achieve the purpose of this study the following questions will be answered:

1. Is there any significant interest in language testing among the language teachers of the Faculty of Engineering?
2. Are there any different comments of language teachers on the appropriateness of the present language tests?
3. Do the language teachers believe that their tests are useful to measure students' success?
4. Are there any cooperation problems among the teachers in the construction of language tests?

5. Are the language tests made up with a clear idea of what they are for?
6. Do the language teachers at the Faculty of Engineering develop their tests according to their specifications?
7. Do the language teachers select the appropriate test items in the construction of their tests?
8. Are the test results interpreted appropriately?
9. What is the way in which language is treated on the present tests?
10. What are the academic qualities of the present tests?
11. Whether some miscellaneous aspects of the test are taken into consideration or not?

1.4. Importance of the Study

It is certain that recent years have witnessed a number of important developments in English language testing. Almost any language-instruction program requires the preparation and administration of language tests. And it can be said that a good testing program resulting in a student's knowledge that he/she is making progress toward clearly specified goals will be a positive motivating force for students and teachers alike.

Therefore, it is more important than ever that tests should encourage good teaching practice.

Obviously, we who teach English as a foreign language have a responsibility to find out how well we have done our tasks or how much improvement our students have made. Because it is clear that testing is an important part of every teaching and learning experience. It contributes directly to the teaching-learning process used in classroom instruction, and it is useful in programmed instruction, curriculum development, marking, reporting, guidance and counselling school administration and research. But effective testing requires careful planning. It is just too easy for any language teacher to take a pen in hand and turn out items without much consideration of the balance or adequacy of the resulting test content. Except by chance, such a test will not prove a very valid measure of the specific skills that the testing situation calls for.

A great deal of literature is available on the methodology of foreign language teaching and more and more suitable teaching material is being prepared for foreign learners. That's why a comparable improvement is needed in foreign language testing and evaluation procedures. Because it is certain that language teaching cannot be distinguished from language testing. We need to know how successful our teaching has been or where it has failed.

Therefore, such a study which attempts to demonstrate the extent of agreement and disagreement on the appropriateness of language tests among teachers and the present situation of an administered test will help to establish some suggestions for constructing and administering better language tests.

1.5. Limitations

1. This research is limited to the opinions of 9 English language teachers of the Faculty of Engineering at Bademlik-Extension Campus of Anadolu University.
2. This study is limited with the questions stated in the questionnaire.
3. Although the questions on the questionnaire are devised sensitively, the questionnaire should not be considered exhaustive because some other less significant aspects of the teaching-learning situation may have been overlooked.
4. Since the questions on the administered quizzes are chosen purposely only a midterm test of 1989-90 academic year is taken into consideration to form a comparison basis for this study.

1.6. Assumptions

1. Most of the items of the questionnaire are taken from a study conducted by The Institute for English Language Education of University of Lancaster, Alderson(1980:1). That's why an adequate coverage of relevant points can assumed to be guaranteed.
2. All the teachers who took part in completing the questionnaire are assumed to have given answers with complete frankness.

1.7. Definitions

Commonly used testing terminology will be given below:

Evaluation: According to Gronlund(1976:6) evaluation is a systematic process of determining the extent to which instructional objectives are obtained by the students. About the same issue, Noll and Scannell (1972:15) point out that evaluation tends to emphasize the role of judgement in considering a variety of types of relevant data or information in determining progress toward a goal, the worth of a program, or how to make a decision. Therefore, evaluation may deal with both quantity and quality of the teaching-learning results.

Measurement: As generally used in education, measurement includes consideration of purposes, interpretation and use of results. According to Gronlund(1976:7) measurement takes only quantifiable units into account. On the same parallel, Noll and Scannell(1972:16) define measurement in terms of ideal attributes of quantitateness and constancy of units. According to them, in measurement we generally strive to be as objective as possible and to minimize the role of judgement.

Testing: Brown(1987:219) indicates that a test in plain, ordinary words, is a method of measuring a person's ability or knowledge in a given area. In other words, it can be said that testing is a data gathering process. With the help of collected data, learner, teacher and materials can be scrutinized. And also it should be said that measurement is only a tool. It is a means to an end, not an end in itself.

Test and Quiz: According to Valette(1977:7) the test covers a specific unit of instruction and is announced in advance. The value of the test lies in the completeness with which it examines the material under study. The essence of the quiz is brevity. The value of the quiz lies in its positive effect on student learning and the practice it affords in the art of test taking.

Objective and Subjective Test Items: Valette(1977) defines an objective item as one for which there is

a specific correct response, therefore it is always scored the same way. For example, multiple-choice type questions are objective items. A subjective item is one that does not have a single answer. A short composition may be scored in different ways by different teachers. That's why, it is considered as subjective item.

Formative and Summative Evaluation: Finocchiaro and Sako(1983:299) define formative assesment as the ongoing assessment of student learning in order to plan or revise subsequent activities in an instructional program. They also point out that the summative assessment is an overall estimate of a student's achievement or proficiency at the end of a given period of instruction.

Norm-Referenced and Criterion-Referenced Tests: Valette(1977) indicates that the norm-referenced test compares a student's performance against the performance of other students. The criterion-referenced test indicates whether the student has met predetermined objectives or criteria.

Discrete-Point and Global Testing: According to Finocchiaro and Sako(1983:297) in discrete-point approach the tests are designed to measure single (discrete) elements of language opposed to the measurement of integrated skills such as reading and writing. On the other hand, global language tests measure the

student's ability to understand and use language in context. Dictation test, cloze test and the tests in which the students are presented with a situation and must express themselves are termed as global tests.

Validity: Finocchiaro and Sako(1983) define validity as the degree to which a test measures what is designed to measure. The validity is always specific to the purpose of the test. Different kinds of tests require different evidences of validity. If a test does this it is said to be valid.

Reliability: Valette(1977) points out that reliability refers to the consistency of measurement that is, how consistent test scores or other evaluation results are from one measurement to another. Reliability provides the consistency that makes validity possible and indicates how much confidence we place in our results. According to Valette(1975:53) reliability of a test can be calculated by Kuder-Richardson formula:

$$R = 1 - \frac{M(n-M)}{n \cdot SD^2}$$

M= Mean
n= number of items

Item Difficulty(Facility Value): According to Finocchiaro and Sako(1983:298) facility value is the percentage of correct responses made by students in relation to the total possible correct answers in a test. It indicates the ease or difficulty of any test item. Valette(1977) gives the following formula

to calculate item difficulty:

$$F.V = \frac{R}{N}$$

R= Number of correct answers

N= Number of students taking the test.

Item Discrimination: Valette(1977:57) states that the discrimination index of an item indicates the extent to which the item discriminates between the testees, separating the more able testees from the less able. The index of discrimination (D) tells us where the students who performed well on the whole test tended to do well or badly on each item in the test. The formula to get item discrimination is:

$$D = \frac{U_{R\ 27\%} - L_{R\ 27\%}}{1/2\ N}$$

D= Discrimination Index

U_R = Upper group of students who got the item right.

L_R = Lower group of students who got the item right.

$1/2\ N$ = Half the number of the students who took the test.

The Mode: The mode is simply the most frequent or popular score in the set, and is determined by inspection.

The Arithmetic Mean: The arithmetic mean, or arithmetic average is the sum of the separate scores divided by their number.

The Median: The median is the midpoint of a series of scores when the scores are arranged in order of their size.

Standard Deviation: According to Valette(1977:67) a standard deviation is a measure of distance away from the mean of distribution. It measures the degree to which the group of scores deviates from the mean. A deviation can be expressed by the formula:

$$D = X - M$$

X= The raw score

M= The mean of all scores.

The formula that expresses the standard deviation is:

$$SD = \sqrt{\frac{\sum (X-M)^2}{N}}$$

$\sum (X-M)^2$ = It is the sum of the square of the deviation of each score from the mean.

N = It is the number of test papers.

CHAPTER II

REVIEW OF LITERATURE

2.0. Chapter Overview

This chapter contains six sections. Each section is considered to have a relationship to each other and provides a basis for this study.

In the first section the educational uses of language tests are given to identify the significance of function and nature of language tests.

Second section deals with the approaches in language testing and briefly describes certain trends in language testing that have become particularly salient within the past several years.

Section three includes the classification of language tests to determine the purpose of each type that is considered to be helpful in constructing the language tests.

In the fourth section the characteristics of language tests are stated. The significance of these qualities in preparing and using foreign language tests is also stated in this section.

Section five includes the construction of language tests. In this section, the significance of factors concerned with the planning and the other procedures of language testing is also stated.

The last section, section six, is devoted to the problems in language testing. In this section, the significance of a logical awareness of the problems involved in a testing system is also stated to determine what is feasible in individual situations.

2.1. Educational Uses of Language Tests

It is certain that testing plays an important role in many Aspects of the school program. Heaton (1975:1) emphasizes the role of testing on language teaching and indicates the two fold interrelation towards learning to evaluate student achievement.

On the same parallel, Finocchiaro and Sako(1983:41) state that teaching and testing could be considered two sides of the same coin. And there is little doubt that in order to enable every teacher to make use of many benefits of a testing program, testing must be presented in a manner that can be understood.

By the same token, Oller(1987:55) points out that good teaching and good testing are, or should be, indistinguishable. He also adds that given a richly organized episodic basis and a variety of testing procedures reflecting the wide-ranging uses of language, language tests can be heart and soul of language teaching.

It is worth noting that testing relates to a multitude of factors which concern both the quality of the teaching given and the learning achieved. Additionally, testing is not something that stands apart from teaching and learning. Because it is generally acknowledged that tests are all extensions of the teaching and learning situation. They may have some advantages for both students and teachers. Properly used, language tests provide the following advantages, Lowe(1975:327).

Advantages for the student:

1. Classroom testing most certainly acts as a stimulus to study and real learning. If the student is made to see the correct purpose of testing, he will learn from his returned examination how well he has mastered the materials he was supposed to have learned, and where his weaknesses are.
2. Good testing encourages a kind of self-examination and relearning process as the

student asks himself why certain answers are incorrect.

3. Good testing invokes for the student better methods of organization and preparation towards tests.
4. Good testing helps to build up the student's confidence in the teacher's fairness and impartiality and his desire to help all his students.

Advantages to the teachers:

1. Good testing is an indication of how well the students have learned what the teacher has tried to teach them, and it points out where they require more help.
2. Good testing feeds back to the teacher information concerning the effectiveness of his teaching methods. So it gives a starting point for consolidating the methods that have been most helpful and for revising those that have not given good results.
3. It enables the teacher to give a more accurate evaluation of the student's ability, or grade, or rank, to the administration, to the student himself, and to his parent or sponsor.

In other words, it may be said that tests can benefit students, teachers, and even administrators by confirming progress that has been made and showing how we can best redirect our future efforts. Moreover, good tests can sustain or enhance class morale and aid learning.

According to Carroll(1961:31) the purpose of testing is always to render information to aid in making intelligent decisions about possible courses of action. He also adds that sometimes these decisions affect only the future design or use of the tests themselves, sometimes the decisions have to do with the retention or alteration of courses of training, and most often, the decisions have to do with the management of the educational careers of individuals.

In Oller's explanation(1987:42) any use of language that can be graded or evaluated may be considered as a language test. That's why, it can be said that if the foreign language teachers are really to teach language skills it is essential for them to do classroom testing.

About the same issue Lado(1961:22) indicates that the theory of language testing is based on present linguistic understanding of language and on observations concerning the role of habit in learning a foreign language. Therefore, it can be claimed that there

are three strands in language testing; language, learning and evaluation. In order to view those issues in some perspective, it will be helpful to look very briefly at historical trends in language testing.

2.2. Approaches in Language Testing

According to Spolsky(1972:215) history seems to break itself into three chronological periods. He distinguishes these three main trends as: 1. The pre-scientific or traditional approach. 2. The psychometric-structuralist or modern approach. 3. The integrative-sociolinguistic or post-modern approach.

2.2.1. The Traditional Approach

According to Spolsky(1972:217) the traditional approach is pre-scientific. Because it devotes no attention to such matters as reliability and objectivity and does not use statistical methods. It, therefore, can be called as methods of examining rather than tests.

Spolsky(1972:14) also indicates that the most frequently used types of tests within the traditional approach were translation, essay-writing, testing knowledge of grammar and sometimes oral tests. Therefore, it can be pointed out that the subjective character of

assessment made it difficult to compare results. This raised doubts about the reliability, and, consequently, about the validity of these tests.

2.2.2. The Psychometric-Structuralist Approach

The second period was identified by Spolsky(1972:218) as the "Psychometric-Structuralist" period, from the early 1950's through the late 1960's. According to him this approach is mainly characterized by the conviction that 'testing can be objective, precise, reliable, scientific'. This led to the development of techniques to make traditional tests more reliable, and, on the other hand, to the development of tests with multiple-choice questions.

The link between psychometrics and linguistics is mainly due to Lado(1961:20). Lado and the others during this period placed great emphasis on identifying and isolating those particular elements of the target language that were anticipated, on the basis of contrastive analysis, to pose specific learning problems for the student.

According to Finocchiaro and Sako(1983:263) in teaching, this approach was exemplified by classroom or language laboratory drills on particular target language structures, and in testing, a series of highly

focused items intended to measure, in a direct and unambiguous way, the student's ability or lack of ability to recognize or produce the particular structures involved.

There's consensus that this was a time when contrastive analysis was a thriving discipline, and that structural linguistics and behavioral psychology combined to provide a "scientific" air to language teaching. Likewise, testing focused on specific language elements such as phonological, grammatical, and lexical contrasts between two languages.

2.2.3. The Integrative-Sociolinguistic Approach

According to Spolsky(1972:225) this period is characterized by a widespread turning away from discrete point structuralist approach toward an approach which views language learning (and language use) not as a mathematical sum-of-the-parts but as a dynamic, highly complex process involving psycholinguistic or sociolinguistic variables. These variables include:

1. the specific communicative intent of the speaker or writer (for example, to inform, convince, dissuade, etc. the listener or reader),
2. the social/cultural roles being played by interlocutors,

3. the nature and amount of shared prior knowledge relative to the information being conveyed,
4. the physical conditions under which the communication is taking place (for example, in a large, noisy party, etc.),
5. a number of other physical, psycholological, and societal aspects of the communicative situation.

In this period, which begins in the early 1970's and continues to the present, the growing dissatisfaction with structuralism and behaviorism led to linguistic research on communicative competence and on the contexts of language. In short, testers were realizing that, as Clark(1983:432) points out the whole of the communicative event was considerably greater than sum of its linguistic elements. Therefore, language testing specialists began to explore ways of testing the communicative competence of second language learners.

It may be said that language testing today reflects current interest in teaching genuine communication, but it also reflects earlier concerns for scientifically sound tests. In other words testing during the last century and the early decades of this one was basically intuitive, or subjective and dependent on the personal impressions of teachers. After intuitive stage, testing entered a scientific stage, a time that stressed objective

evaluation by language specialists. And it can be said that we are now in a communicative stage, a time when emphasize evaluation of language use rather than language form.

In the recent literature cloze test, editing test, dictation and communicative tests have received quite a bit of attention. As Oller(1978:27) points out a cloze test is merely a passage with blanks inserted for words that have been deleted. In other words a standard cloze test is a text in which every nth word has been deleted. In cloze test the testee has to fill in the blanks. In an editing test he has to delete what is alien. An editing test is constructed by randomly inserting new words into an existing text. These words are systematically taken from some other text.

According to Heaton(1975:186) dictation as a testing device measures too many different language features to be effective in providing a means of assessing any one particular skill. A standard dictation task is one in which a learner has to write out in full an orally presented text.

Developments in communicative testing can be viewed as attempts to come to terms with the communicative language teaching movement. Brown(1987:230) indicates that a communicative test has to test for grammatical, discourse, sociolinguistic, and illocutionary competence

as well as strategic competence. It can be said that communicative testing has some clear advantages: the behaviour-oriented character, the specification of language use, and the close relation to the needs of the learner. However, there are no clear results of empirical research into communicative testing.

With this brief historical back-drop, we can now look at some of the ways tests can be classified.

2.3. Classification of Language Tests

According to Harrison(1983:4) the appropriateness of a test is largely determined by its purpose. Language tests may be used for different purposes and there are many kinds of tests.

On the same parallel, Spolsky(1972:28) distinguishes foreign language tests into two classes according to the purposes for which they are used. According to him in the first class there are achievement tests, those concerned with what has been taught, and diagnostic tests, those concerned with what is about to be taught. In the second class, there are achievement tests, concerned with what the subject can do now and predictive tests concerned with what he should be able to do in the future. On the other hand, he also states that this temporal distinction is less important than the major functional one; exactly the same test can serve

as a diagnostic test before some material is taught and as an achievement test after it. Similarly, proficiency tests are generally used as predictors of future performance.

According to Davies(1978:45) language tests may be distinguished according to their four uses. These uses are achievement, proficiency, aptitude and diagnostic. He also symbolizes the differences in the following figure 1.

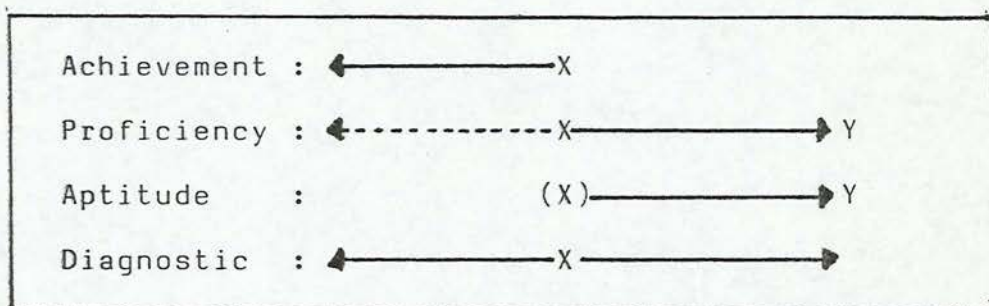


Figure 1. The Differences in the Four Uses of Language Tests. Davies(1978:45).

As it is seen in the figure 1, the four uses are distinguished firstly in terms of subject matter. Thus, a left -pointing arrow represents past time and a right- pointing arrow represents future time. In this figure "X" is the point at which the test is taken. Additionally, there is a broken arrow which represents what has been learned of a known or unknown syllabus. It can be also seen in the figure 1 that the arrow relates the language test under proficiency to something else 'Y', and under aptitude to itself(X).

2.3.1. Achievement or Attainment Tests

According to Davies(1978) these tests are concerned with assessing what has been learned of a known syllabus. And he also indicates that although the primary interest in the past, very often some further use is made of the same test in order to make meaningful decisions about the pupils' future.

In Finocchiaro and Sako's explanation(1983:15), these tests are designed solely to cover material that has been presented in the classroom (or in a laboratory) during a period, but not the material which may not have been taught. They also state that achievement tests may be subdivided according to the time of the test administration and the scope of the material. Language achievement tests may be characterized as objective, subjective, or semi-objective according to the degree of possible bias in scoring or grading.

Finocchiaro and Sako 1983:16) defines some subtypes within the achievement test category. These are: Language behaviour tests, Specialized Terminology Test, Criterion-Referenced Tests, Cloze Tests and dictations. According to them, language behaviour tests are designed to determine the listening or reading comprehension of the individual through the direct observation of his/her behaviour or performance. And the Specialized Terminology Tests are constructed

primarily to measure the learner's understanding or production of expressions peculiar to a technical or professional field of study.

2.3.2. Proficiency Tests

According to Davies(1978:46) proficiency tests are concerned with assessing what has been learned of a known or unknown syllabus, hence the broken arrow in Figure 1. In other words, the aim of the proficiency test is to determine whether this language ability corresponds to specific language requirements.

Finocchiaro and Sako(1983:21) states that language proficiency tests are designed to measure control of language or cultural items and communication skills already present at the time of testing, irrespective of formal training. Such tests are generally used for specific purposes, for example, to determine selection of students for a specific program, to ascertain appropriate placement levels in courses or in advanced training programs, and to judge the examinee's readiness to perform specific tasks in a work activity.

On the other hand, proficiency tests may also be classified according to the degree of bias of scoring. Many proficiency tests are considered as standardized tests. Finocchiaro and Sako(1983) give the following types as the examples of language proficiency tests:

Michigan Test of English Language Proficiency, Common Concepts Foreign Language Tests, English Comprehension Level Examinations, TOEFL Tests, The Pimsleur Modern Foreign Language Proficiency Tests, The Cambridge English Examinations, The MLA Foreign Language Proficiency Tests for Teachers, and many others.

2.3.3. Aptitude Tests

According to Davies(1978) aptitude tests assess proficiency in language for language use. Therefore, in Figure 1 the arrow relates the language test under proficiency to something else (Y), and under aptitude to itself (X). He also adds that an aptitude test is generally conceived of as assessing the amount of linguistic skill needed for learning languages. By the same token, he represents the aptitude test as (X), in brackets in his figure. Because in aptitude research the goals should be set up first.

As Finocchiaro and Sako(1983:23) point out language aptitude tests (sometimes termed prognostic or predictive tests) are designed to predict, before beginning language study, a subject's probable capability of acquiring the language. According to them, aptitude test items center about factors considered predictive by researchers such as sound discrimination, association of sounds and symbols, grammatical sensitivity, and

language interest (motivation for learning). Examples of aptitude tests are the Modern Language Aptitude Test (that is, the Carroll and Sapon Test), the Defense Language Aptitude Test, the Pimsleur Language Aptitude Battery, and the English Language Aptitude Test.

2.3.4. Diagnostic Language Tests

Davies(1978:47) believes that the diagnostic use differs from the others in that it relates to the use of the information obtained and to the absence of a skill in the learner. Achievement, proficiency and aptitude are all concerned with both use and skill. A diagnostic test is a use made by a teacher of the information provided from the presence or absence of part of one of the skills.

According to Finocchiaro and Sako(1983:22) a diagnostic language test seeks to identify the specific strengths and weaknesses of a foreign language student. A very specific weakness such as misuse of verb forms in several test items may be revealed, or more general problems in integrated skills of listening, reading, speaking, or writing may be brought to light.

Finocchiaro and Sako(1983) also indicate that a well-known example of this type of test is the Diagnostic Test for Students of English as a Second Language,

which is designed to give part scores in the specific areas of English structure and idiomatic vocabulary.

2.4. Characteristics of Tests

It is certain that a teacher who is unaware of the relationship between the content of a test and the consistency of its results is in danger of writing tests which produce misleading information about his students. Because certain criteria should be satisfied in an effective testing process.

As Krashen(1982:176) points out a good test needs to meet certain standards. In other words, a good test must be (1) valid, (2) reliable and (3) practical. These three principal qualities are generally considered when developing and judging language tests. There is little doubt that the more valid, reliable and practical a test is, the better the evaluation is.

These characteristics can be divided into other subgroups, but they are large enough to cover the main features involved on an ideal test. That's why a brief review of these criteria will help us to see their place and requirements in a testing situation.

2.4.1. Validity

Test validity is considered as the most critical factor to be judged in the total program of foreign

language testing. Heaton(1975:153) points out that the validity of a test is the extent to which it measures what it is supposed to measure and nothing else.

About the same issue, Lado(1961:30) indicates that validity is not general but specific. He states that if a test of pronunciation measures pronunciation and nothing else, it is a valid test of pronunciation. It would not be a valid test of grammar or vocabulary because it does not test grammar or vocabulary.

According to Finocchiaro and Sako(1983:24) two questions must be considered when determining test validity of a foreign language test. What aspects of the language is the test designed to measure, and how well does it, in fact, measure the global skills or the discrete elements of the language.

There are many different kinds of validity. Below are short explanations of five primary validity concepts.

a. Content Validity: Harrison(1983:11) states that content validity is concerned with what goes on into the test. It is assured by checking all items in the test to make certain that they correspond to the instructional objectives of the course, whether they are discrete or integrative language skills.

According to Finocchiaro and Sako(1983:25) given the measurement objective of the test(diagnostic,

achievement, etc.) the closer the test items correspond to the functional or significant elements of the language, the higher the content validity.

b. **Concurrent Validity:** In validating a new test, the writer seeks to obtain a high correlation of test scores with an independent criterion or standard. According to Finocchiaro and Sako(1983:26) a mechanism for ascertaining concurrent validity could follow a pattern such as the following: A new language test is administered to students in the course for which the test is developed, and scores are recorded for each student. These scores are then compared to the criterion test grades or the teachers' ratings. The relationship of the two is a measure of concurrent validity.

c. **Predictive Validity:** This type of validity is especially used in the validation of language aptitude tests. Finocchiaro and Sako(1983) point out that this validity is generally determined by administering the test to a group of prospective students, following their progress in their language courses, obtaining data about each of them on some criterion measure and then computing the correlation coefficient between the aptitude test scores and the criterion grades or data. The higher the correlation, the better the predictive power of the test.

d. Construct Validity: This validity measures specific characteristics expected of the student in certain situations. First those specific characteristics should be presumed and then be tested. If test results confirm them we can say that the test has construct validity.

According to Finocchiaro and Sako(1983) construct validity indicates why a test may be useful(what trait it really measures) and whether the test might also be useful for other purposes when knowledge about the demonstrably measured trait is important.

e. Face Validity: It is concerned with what teachers and students think of the test. Face validity is considered as a judgement about a test based on the way the test looks to educators, students, and the general public. Finocchiaro and Sako(1983) indicate that the test should not only "be right", it should also "look right".

According to Harrison(1983) the only way to find out about face validity is to ask the teachers and students concerned for their opinions, either formally by means of a questionnaire or informally by discussion in class or staff room.

2.4.2. Reliability

It is generally acknowledged that reliability is a prerequisite criterion for a valid test. The

reliability of a test is considered as its consistency. According to Paterno(1961:20) a test that lacks reliability is as useless as a thermometer that gave different readings when the temperature of the air was the same.

Brown(1987:221) indicates that if we give the same test to the same subject or matched subjects on two different occasions, the test itself should yield similar results; it should have test reliability. In Oller's explanation(1979:4) the reliability of a test is a matter of how consistently it produces similar results on different occasions under similar circumstances.

Heaton(1975:155) defines the term "reliability" in great detail and says that reliability is of primary importance in the use of achievement, proficiency and classroom tests. According to him, it is worth noting that a test can be reliable without necessarily possessing validity. However, it can be said that reliability is clearly inadequate by itself if a test does not succeed in measuring what is supposed to measure.

Finocchiaro and Sako(1983:28) point out the following factors which may affect the reliability or stability of a test: variation in testing conditions, differences in administrative instructions, inaccuracy in scoring, inadequate sampling of test items and

reducing subjective scoring of the test. These factors will tend to decrease test reliability.

2.4.3. Practicality

It can be said that the main questions of practicality are administrative. Practicality is a general term to cover some criteria of usability, such as economy, scorability and administrability of tests.

According to Finocchiaro and Sako(1983:30) one valid step in reducing the cost of testing is through the reuse of test booklets. Tests should be as economical as possible in cost. The practicality of a language test is further determined by its ease of scoring. Scoring which can be done by hand through the use of scoring keys or by machines provides objectivity in test grading. Another factor in test practicality is that the testing instrument must be easy to administer to the examinees. Instructions of the tests should be clear enough to the students. If the students fail to understand the procedure, they cannot demonstrate their actual performance.

Tuckman(1975:302) introduces the following checklist to determine the usability of a test. If a test satisfies these aspects it is a practical test.

Is My Test Usable?

1. Is it short enough to avoid being tedious?
 - a. Does it stop short of creating fatigue, stress or boredom?
 - b. Have I tried to make it as short as possible within the limits of reliability?
2. Is it practical for classroom use?
 - a. Can it be used conveniently in a classroom?
 - b. Is it within the limits of available teacher time?
 - c. Can it be used to test all students?
 - d. Is it realistic about the kinds of equipment and physical set-up it requires?
3. Are there standard procedures for administration?
 - a. Are there clear, written instructions?
 - b. Can it be administered by someone other than me?
 - c. Can it be given in a nonthreatening, nondiscriminatory way?
4. Can students comprehend it and relate to it?
 - a. Is it written at a level students can understand?
 - b. Is it interesting, clever, or provocative?
 - c. Is it written to engage students?

It is certain that to prepare valid, reliable and practical classroom tests is greatly achieved if a series of basic steps is followed. In the following section the basic steps of classroom testing will

be presented to determine the features of a feasible test and to analyze an administered test at the Faculty of Engineering.

2.5. Construction of Language Tests

According to Harris(1961:27) the construction of an educational test includes the following steps':

1. Planning the test.
2. Preparing the test items and directions.
3. Submitting the test material to review and revising on the basis of review.
4. Pretesting the material and analyzing the results.
5. Assembling the final form of the test.
6. Reproducing the test.

On the same parallel, Gronlund(1976:35) introduces the following Figure 2. to describe basic steps in classroom testing.

According to Gronlund(1976) the first four steps in this figure are the factors concerned with the planning of classroom tests. Steps 5-8 are the factors concerned with the other procedures in testing.

We have chosen to describe the planning of an achievement test in this study. Because the tests we administer at the Faculty of Engineering, can be classified in this group. By the same token, it is

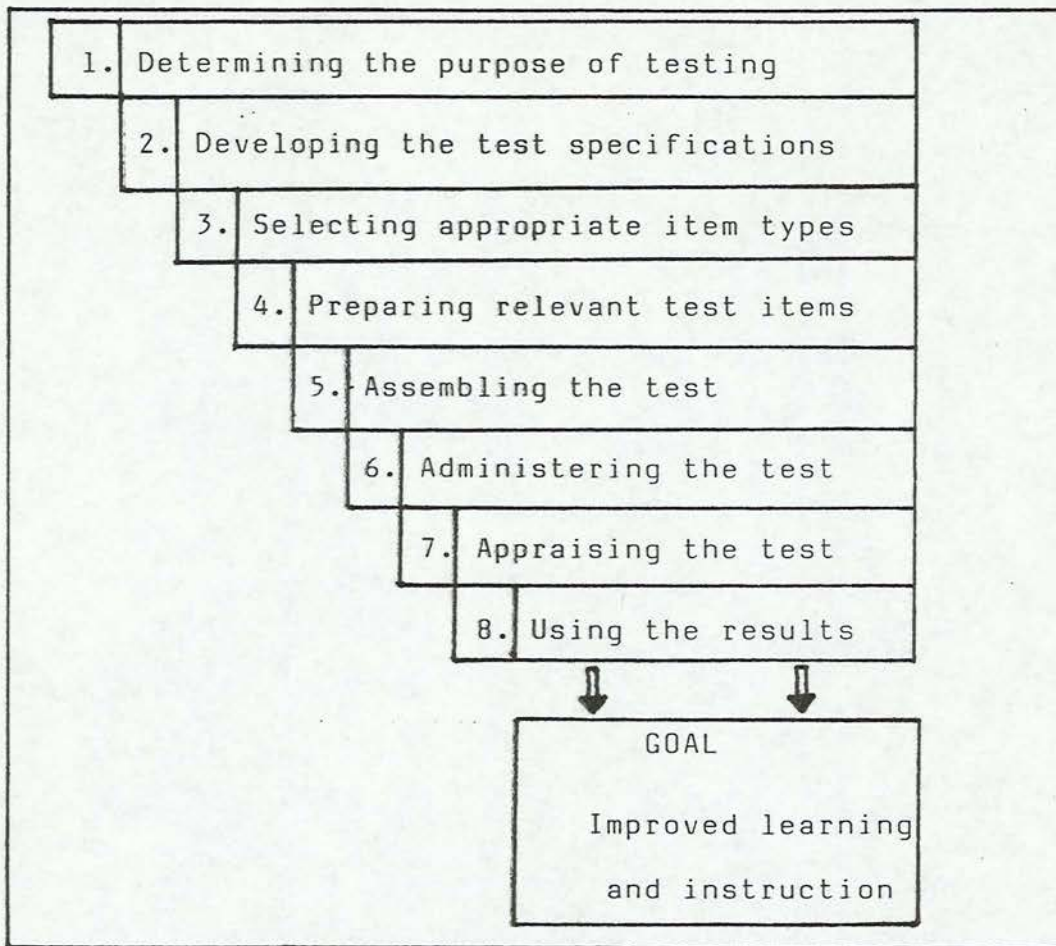


Figure 2. Basic Steps in Classroom Testing,
Gronlund(1976:35)

certain that very similar procedures would be followed in the planning of a general proficiency test, except that the test outline would have to be built on a broader base than simply the content of a single course.

2.5.1. Determining the Purpose of Testing

Harris(1961) indicates that in the preparation of an achievement test, one may base the test objectives directly on the objectives of the course. Achievement

tests should directly utilize the vocabulary, grammatical structures and content of the texts and materials which have been used prior to testing.

As Finocchiaro and Sako(1983:97) point out objectives naturally differ with the type of language test to be constructed, the age and learning level of the student, and other such relevant factors as the country or school-wide policy, the curriculum, and the importance given to foreign languages in the nation as a whole. That's why, if the task is to construct a language achievement test, the objectives should adhere closely to the instructional material used and to the course content.

2.5.2. Developing The Test Specifications

Gronlund(1976:54) cites the following questions as criteria to develop sound objectives for a particular course.

1. Do the objectives include all important outcomes of the course?
2. Are the objectives in harmony with the general goals of the course?
3. Are the objectives in harmony with sound principles of learning?

4. Are the objectives realistic in terms of the abilities of the students and the time and facilities available?
5. Are the objectives defined in terms of changes in pupil behaviour?

The test specifications can be developed by obtaining a representative sample of student performance in each of the areas to be measured. According to Harris(1961) the first step is determining the general course objectives. General instructional objectives are formulated by the national curriculum authorities. For the teachers who teach the same subject at the same school the statements from the national curriculum which are based on a taxonomy of educational objectives serve as a frame of reference. Because the national curriculum center, of course, does not state objectives for every school and for different text books. Therefore, the teachers study the contents of the text book and the objectives in the given curriculum in relation to their particular situation.

Harris(1961) defines the next step as dividing the general course objectives into their components. Because the objectives we defined in step one were extremely broad. In this step, listening, speaking, reading, and writing each include four elements. These are:

1. Phonology/orthography: The sound system (in listening and speaking) and graphic system (in reading and writing).
2. Grammatical structure: This is the system of grammatical signaling devices. Not all grammatical patterns are equally common in the written and the spoken forms of the language.
3. Vocabulary: The lexical items needed to function effectively in each of the four skills.
4. Rate and general fluency: The speed and ease with which the user of a language can decode and encode messages.

2.5.3. Selecting The Appropriate Item Types

It can be said that the items used in classroom tests are typically divided into two general categories:

1. The objective item, which is highly structured and requires the student to supply a word or two, or to select the correct answer from among a limited number of choices.
2. The essay question which permits the students to select, organize, and present the answer in essay form. Selection should be made according to the purpose of testing.

According to Heaton(1975:13) a good classroom test will probably contain both subjective and objective test items. And he presents following examples of objective test items (excluding multiple-choice items):

1. Transformation:

John Fairbanks is a very fast runner.

John Fairbanks runs

2. Completion:

The time is now half ten

3. Combination:

Steve did his homework. Then he went swimming.

(After)

4. Addition (Grammar):

YET Haven't you seen this film?

5. Rearrangement:

at/poor/Look/that/girl/young

6. Correct/Incorrect (or True/False)

Put a tick if the statement is correct and a cross if it is incorrect.

1. The sun rises in the East ()

2. Fish can't fly, but birds can ()

3. When ice melts, it turns into water ()

Heaton(1975) also adds that although the examples have been confined to the testing of grammar, many of the item types can be used equally, successfully

to test vocabulary as well as aspects of the reading, writing, listening and speaking skills.

It is obvious that each type of test item is efficient for measuring some learning outcomes and inefficient, or inappropriate for measuring others. The multiple-choice item is the most generally adaptable type of test item. The essay and other supply types are the most effective for measuring the ability to organize data, the ability to present original ideas, and for some types of problem solving situations.

2.5.4. Preparing Relevant Test Items

Finocchiaro and Sako(1983:64) state that only teacher can decide which items fulfill students' needs best. It is obvious that some test items and tasks will test the students' recognition of some language elements, while others will measure their ability to produce the element in a meaningful context.

Therefore, in attempting to construct test items of average difficulty we should avoid resorting to undesirable methods for obtaining difficulty. For obvious reasons the following barriers should be eliminated from the test.

1. Ambiguous statements
2. Excessive wordiness

3. Difficult vocabulary
4. Complex sentence structure
5. Unclear instructions
6. Unclear Illustrative material

On the other hand, some clues lead the nonachiever to the correct answer and thereby prevent the items from functioning as intended. Some common clues in test items are these:

1. Grammatical inconsistencies
2. Verbal associations
3. Specific determiners
4. Phrasing of correct responses
5. Length of correct responses
6. Location of correct responses

2.5.5. Assembling The Test

It can be said that no matter how carefully we prepare test items, our test still needs reviewing. According to Harris(1961:29) in the preparation of multiple-choice or short answer(supply) items, it is always necessary to write more items than will be needed in the final form of the test. Because some items can easily be added or deleted after a careful review of the complete collection of items.

As Finocchiaro and Sako(1983:207) point out when all the test items have been written, they should

be examined or edited by a fellow language teacher or a language specialist. As we attempt to increase the difficulty of an item, we may introduce some ambiguity. These problems may include: no clear-cut answer, double answers, rewording needed for clarity, testing objectives not apparent, improvement needed in distracters, or inadequate variation in the order of distracters from group to group.

On the other hand, written directions and information are considered as the necessary part of the test. According to Harris(1961) test directions should be brief, simple to understand, and free from possible ambiguities. The following information should be put on the test paper:

1. Name of the subject.
2. Content of the test: The content area of the test is an aid to classify the items for the item bank. It is also necessary for content validity.
3. Number of items and scores allotted to each item should be written.
4. Time allowed for answering: A practical solution to this is that; after the test is assembled an English teacher answers all the questions, the time he spent is doubled for the students.

5. Basis for answering: The directions for each section of the test should indicate the basis for selecting or supplying the answer. To include sample test items is helpful for the students.
6. How to mark the answers: Directions should indicate whether the students circle, underline, put a tick or put a cross out the correct item.
7. Page notice: Experience has shown that some students have missed the items on the back pages. Notices like, "turn over", "continue to the next page", and "end of test" are useful.
8. Directions for guessing: If there is a correction for guessing, the directions might say: "If you do not know the answer, do not guess, for guessing will lower your score." If there is no penalty for guessing the directions must say: "Try to answer every question, even if you are not sure your answer is correct".

2.5.6. Administering The Test

According to Finocchiaro and Sako(1983:217) test administration problems include the control of

testing environment, test examiner qualifications, the presentation of test instructions, the timing of the test and collection of test materials at the end of the test.

They also point out that an ideal testing environment may be hard to obtain, but attention should be given to the following factors:

1. The testing room should be quiet and well illuminated.
2. Ventilation, temperature, and humidity should be controlled when possible.
3. Seats should be far enough apart to prevent distraction or coping of information from neighboring examinees.

On the other hand, the examiners should also be familiar with the purpose of the test, the test materials, and the directions for the examinees. The examiner must tell the examinees exactly what they are to do.

It is generally acknowledged that in order to observe uniform testing conditions for all testing sessions, time must be accurately adhered to.

Additionally, it is certain that scoring procedures for language tests are very important. Because scoring some items causes some difficulties and differences

in scores at our schools. As Finocchiaro and Sako (1983:219) define there are two principal types of scoring: objective and subjective.

Objective scoring: When a test can be scored with no variation of opinion among the possible scorers, the scoring of such a test is said to be objective. In objective scoring tests should be scored by personnel who have had training and experience in scoring, and, for many types of tests, by persons who know both the native and the foreign language.

Subjective scoring: The evaluation of overall oral proficiency is necessarily a subjective type of measurement. According to Finocchiaro and Sako (1983:221) the scoring of this type of test requires more highly skilled persons than those needed for scoring objective-type tests. In evaluation of writing proficiency, a test that requires only the production of words is the easiest to grade and is the least subjective while a test that requires production of sentences and paragraphs is the most difficult to grade and is also the most subjective.

Finocchiaro and Sako(1983) introduce one example of weights given in the evaluation of essay test components in the following figure 3.

According to Finocchiaro and Sako(1983) in figure 3, the weights are determined for each of the

Weight in Percent	C o m p o n e n t s
22	Quantity of writing produced (number of valid lines or paragraphs)
20	Number of valid ideas and their organization into a logical, coherent, comprehensible whole
22	Grammatical correctness and appropriateness
22	Choice (richness and appropriateness) of vocabulary and idiomatic expression
5	Sensitivity to the situational context (where this is pertinent)
2	Punctuation
2	Spelling
5	Style (appropriateness, coherence)
0	Penmanship

Figure 3. An Example of Evaluation of Essay Test Components, Finocchiaro and Sako (1983:231).

elements involved in an essay test. They also indicate that to minimize the subjectivity of evaluation, model answers should be prepared at each learning level in which an essay is given. It is worth noting that although the Figure 3 represents an example of evaluation of essay test components, the weights in any given test may vary according to the communicative, social, or professional needs of the students. That's why,

it can be said that weights should be determined in advance and preferably by examiners who have had thorough training in rating similar tests.

2.5.7. Using The Result

According to Davies(1978:42) and Finocchiaro and Sako(1983:32) evaluation results are used for the following purposes:

1. Progress: The point of tests with a progress aim is to compare pupils not with one another but with an already determined standard. To this end, these test results can be used as diagnostic measures to indicate weaknesses and strengths to both the teacher and the student and, where appropriate, the parent. In other words, evaluation results can be used to report student progress to teachers, parents, etc.
2. To use in guidance and counselling: Evaluation results are used in helping the students solve personal and social adjustment problems all require an objective knowledge of the student's abilities, attitudes and other personal characteristics.
3. To use in school administration: Collected data from evaluation helps to judge the

extent to which the objectives of the school are being achieved, to identify strengths and weaknesses in the curriculum. Evaluation also provides the information on which to base administrative decisions concerning the placement, grouping and promotion of students.

2.6. Problems In Language Testing

As Harris(1961:312) points out in the last twenty-five years, the methods used in the teaching of English as a foreign language-or for that matter, the teaching of any language as a foreign, or second, language-have been modified, expanded, improved, examined, and dicussed. He also states that unfortunately, many classroom teachers are still left with the problem of developing materials for use with their students. Moreover, even if they are able to obtain a useful series of integrated texts, these teachers are still faced with the problem of writing supplementary material to fit their particular needs as well as preparing language tests.

A large number of teaching aids have proved to be very useful in second-language instructions. Introduction of challenging teaching aids enables people to learn a foreign language in such a short

time interval which could not have been dreamt of in the 1950's or 60's. Naturally this rapid progress has brought new perspectives into language testing, and some traditional aspects of tests have been revised. Concepts such as communicative competence and real-life language skills have changed the designs of tests.

According to Madsen(1983:7) in this communicative era of testing, the best exams are those that combine various subskills as we do when exchanging ideas orally or in writing. That's why, now teachers sometimes do not feel themselves comfortable with deciding the appropriateness of tests by which they measure student assessment.

It can be said that there are some technical problems dealing with appropriateness of tests. According to Kohonen(1976:34) if we want to measure our students' ability to express themselves in a second language, there is probably no other way for doing this properly than having them talk and than assessing the performance subjectively, no matter how time consuming or unreliable such measurement may be.

On the other hand, one of the most outstanding shortcomings of modern language testing is the insufficiency of tests of production. According to Heaton(1975:7) production items do not always guarantee that the student will deal with the specific matter

the examiner had in mind. In this case the test item is not entirely satisfactory.

There are several unanswered questions about objective tests. The problem of giving more emphasis on the negative aspects of language learning has not been solved by many testing techniques.

Briefly, it may be said that there is no single solution to problems of second language testing mentioned in this part. As Kohonen(1976:41) proposes that problems have to be solved case by case, depending on the purpose of testing: what kind of second language command a teacher wants to measure, and what he needs the information for. Thus, the choice of the types of tests and items to be used must be considered in the light of current aims of teaching, purpose of testing, and resources available.

According to Henning(1982:33) there are some common testing mistakes for EFL teachers to avoid. As he points out the lists of common testing problems provided here can be grouped into four categories as follows: General examination characteristics, item characteristics, test validity concerns, and administrative and scoring issues.

General examination characteristics:

1. Tests which are too difficult or too easy:
When tests are too difficult or too easy,

there is an accumulation of scores at the lower or higher ends of the scoring range. So as a result of such boundary effects, there is information loss and reduced capacity of the test to discriminate among students in their ability.

2. An insufficient number of items: Tests may be too long and thus needlessly tire the students. On the other hand, a more common mistake is for a test to be too short and thus unreliable.
3. Redundancy of test type: In testing general language proficiency, it is common practice to devise a battery of subtests to ensure that all important language skills are covered by the test as a whole. It has been demonstrated, for example, that inclusion of subtests of error identification, grammar accuracy, vocabulary recognition, and composition writing "leaves no room" for a subtest of listening comprehension. Many such indiscriminately maintained proficiency tests are inefficient in the sense that they carry too much extra baggage.
4. Lack of confidence measures: When logically developed tests are used for important evaluative decisions, estimates of reliability and

validity should be provided for these tests. It is not uncommon for unwarranted reliance to be placed in some foreign standardized test when the characteristics of the population with reference to which it was developed are vastly different from those of the population with which it is being used. Vast differences of this sort imply a need for reanalysis of the test in the new situation.

5. Negative washback through non-occurrent forms: Through use of inappropriate structures of the language it is possible to teach errors to the students. Consider the following items:

I here since two o'clock.

a. am being b. have been c. will be d. am be

Option 'd' clearly does not exist in any natural context in the English language. The possibility exists that a learner, particularly at the beginning stage, might learn this form and entertain the thought that 'am' may serve as an auxiliary of be.

Item characteristics:

6. Trick questions: The use of trick questions must be avoided. Consider the following example:

I did not observe him not failing to do his work because he was

- a. always working.
- b. ever conscientious.
- c. consistently lazy
- d. never irresponsible

Here, the stem contains a double-negative structure that stretches the bounds of normal English usage. That's why, many of the better students who have greater mastery of the lexicon can be fooled, while weaker students manage to pass, perhaps by attending to the fact that option 'c' is different from the other options.

7. Redundant wording: A common problem in item writing, particularly of multiple-choice items, is needless repetition. An example would be the following:

He went to school

- a. because he wanted to learn more.
- b. because he wanted to meet new friends.
- c. because he wanted to get a better job.
- d. because he wanted to please his parents.

Such an item is better written as follows:

He went to school because he wanted to

- a. learn more.
- b. meet new friends.

- c. get a better job.
- d. please his parents.

8. Divergence cues: In writing options for multiple-choice type items it is important not to provide cues regarding the choice of the correct option. Consider the following example:

In the story it was raining because

- a. the sky was dark.
- b. everyone insisted on wearing an overcoat and carrying an umbrella outside.
- c. it was that time of the year.
- d. they heard thunder.

Without having read the story, we would imagine that 'b' was the correct answer merely because greater detail is offered.

Test validity concerns:

9. Mixed content: Sometimes tests have been claimed to measure something different from what many of their items are actually measuring. The following two items are offered by way of example. The first item was said to measure knowledge of verb tenses; the second was said to measure vocabulary recognition:

He the man yesterday.

- a. see
- b. saw
- c. will see
- d. is seeing

The lady to many cities in Europe last year.

a. visited b. traveled c. visits d. climbed

In the first example, purported to test tense, we find option 'a' actually measures knowledge of subject-verb agreement. Similarly, the second item, supposedly measuring vocabulary recognition, includes option 'c' which test tense. These kind of inconsistencies make for invalid tests.

10. Wrong medium: Sometimes reading-comprehension questions require accurate written responses to show comprehension of the passage. According to Henning(1975:27) such tests are invalid in the sense that they measure something other than what they are intended to measure. Care must be taken that the response medium be representative of the skill being tested.

11. Common knowledge: Items that require common-knowledge responses must also be avoided. Consider the following items:

According to the story, Napoleon was born in

a. England b. Germany c. France d. Italy

Obviously, responding correctly to such an item does not entail the ability to comprehend a reading passage.

Henning(1982:37) also states that these problems are surprisingly common in the preparation of classroom EFL tests. He adds that other problems may also be cited, but the ones given here are certainly among the more common.

Grittner(1969:342) introduces the following checklist of basic principles for constructing achievement tests to serve as a guideline for EFL teachers in the preparation of their own examinations.

A checklist of basic principles for constructing achievement tests:

1. Test that which has been taught. Normally tests can involve:
 - a. Simple recall of prelearned material.
 - b. Rearrangement of prelearned material.
2. In presenting test items it is considered good practice to:
 - a. Begin with the simplest items.
 - b. Proceed to the more difficult items.
3. In selecting test items remember to:
 - a. Choose problem sounds, structures, and word-order sequences.
 - b. Avoid tongue twisters and absurdities.
4. In setting up multiple-choice items it is common practice to:

- a. Include at least one 'distractor'.
 - b. Include two plausible but incorrect items.
 - c. Include only one unambiguous item which correctly answers the question.
5. Directions on the test should be in English. However, the questions (multiple-choice items) should be mostly in the target language.

To sum up, testing plays an important role in many aspects of the school program. Because tests are considered as all extensions of the teaching and learning situation. And also, historical trends reflect that the parallel between language teaching and testing represents a healthy correlation between the two fields. That's why, it can be said that there are many kinds of tests, each with a specific purpose and a particular criterion to be measured. Additionally, it is generally acknowledged that the three main requirements of a test are validity, reliability and practicality. If these three axiomatic criteria are carefully met, a test should then be administrable within given constraints, be dependable and actually measure what it intends to measure. Thus, it is obvious that anyone designing language tests must have a clear idea-or as clear an idea as possible- of both why he is testing and what he is testing. And if the common testing mistakes are avoided or resolved, the quality of EFL testing will improve.

CHAPTER III

THE RESEARCH METHOD

In this chapter "Research design", "Subjects", "Data collection", and "Data analysis" are presented.

3.1. Research Design

In order to answer the questions introduced in the "Problem" section (cf. 1.2) and to achieve the purpose set in the "Purpose of the study" section (cf. 1.3), an extensive questionnaire was prepared to demonstrate the extent of agreement or disagreement on the appropriateness of English language tests among teachers (cf. Appendix A-1).

The questions of the questionnaire are intended to reflect the updated opinions in the field of testing in a global form. Therefore, the questionnaire can

serve to provide innovations for the teachers who know little about testing. For the subjects' ease the questionnaire was introduced them in the form of an interview and the answers were recorded by the interviewer. The underlying purpose of this technique was to minimize the mechanical nature of the questionnaires and to make the teachers feel comfortable.

In this study, there is also an analysis of a midterm test administered at the Faculty of Engineering in 1989-90 academic year (cf. Appendix B-1). This test was administered to the students soon afterwards of the questionnaire applied to the teachers. In the analysis of the test, the testing mistakes were classified into three categories according to the checklists given by Henning (1982: 33) and Grittner (1969: 342), (cf. 2.6).

3.1.1. Description of Questionnaire

In this study, the questionnaire includes twenty four questions. The questions of the questionnaire can be classified in eight groups according to their varying aspects.

In the first group, there are three questions (Items 1-3). These questions are about teaching and testing experiences of the subjects. Second group includes two questions (Items 4-5) to collect data about present tests and objective testing.

Two questions (Items 6-7) can be considered in the third group to ascertain information about the comments of teachers on the certain types of tests and purposes of language tests. In the fourth group, there are two questions (Items 8-9) about the extent of cooperation between the teachers at the Faculty of Engineering.

Group five includes three questions (Items 10-12). This group aims at collecting data about the opinions of teachers on the use of test results. In the sixth group, there are four questions (Items 13-16) on the language learning issue related with the tests.

In the seventh group, there are two questions (Items 17-18). These questions concentrate on the opinions of teachers about the way in which learner is taken into account on tests. Finally, the last group, group eight, includes five questions (Items 19-23). These questions deal with the opinions of teachers about miscellaneous aspects of the tests.

3.2. Subjects

In this study, nine English language teachers of the Faculty of Engineering at Bademlik-Extension Campus of Anadolu University comprised the interviewees of the questionnaire.

There are nine teachers of English as a foreign language at the Faculty of Engineering. The number of teachers who have a length of experience between 5 and 9 years is six. One teacher has 19 years of experience and two most experienced teachers have been teaching for thirty years. All of them served as the subjects for this study.

3.3. Data Collection

In this study, data was collected through a questionnaire and by an analysis of a midterm test which are supposed to be suitable for the purpose of the study.

In order to collect data, the results of the questionnaire were interpreted and evaluated through the help of tables for each question.

On the other hand, in the analysis of the sample test three categories are taken into consideration. The general examination characteristics, item characteristics and test validity concerns of the sample test are evaluated. Data collected from the analysis of the midterm test was contrasted with the results taken from the questionnaire in the Discussion and Suggestions Chapter (cf. 5.1).

3.4. Data Analysis

The results of the questionnaire are made significant through the application of percentage system. In the analysis of the results 23 tables were used. This means that there is one table for each question.

Although a closed questioning technique was chosen there are some open-ended questions in the questionnaire. Unfortunately, the subjects didn't pay any attention to those open-ended questions. Therefore, I found this limitation a serious handicap in determining the tendency to help to improve the present language testing system among classroom teachers.

CHAPTER IV

ANALYSIS OF RESULTS AND DISCUSSION

The specific aim of this study is to demonstrate the extent of agreement and disagreement on the appropriateness of language tests administered at the Faculty of Engineering, Anadolu University, from the language teachers' points of view. This study also tries to find out information about the present situation of English language tests administered at this faculty. To achieve these tasks, the research has been carried out by use of a detailed questionnaire and by an analysis of a sample test administered at this school.

4.1. Evaluation and Interpretation of Results

This part consists of the analysis of teacher's responses to the questions presented in the questionnaire.

The questionnaire includes twenty four questions. The total number of interviewed teachers is 9. The results of the questionnaire will be interpreted and evaluated through the help of tables for each question.

TABLE 1

Teaching Experiences of the Teachers

Years in Teaching	Number of Teachers
5- 9	6
15-19	1
20-30	2

As can be observed in Table 1, most of the teachers have 5-9 years of teaching experience. One teacher has 19 years of experience and two most experienced teachers have been teaching for thirty years. This indicates that the teaching staff at the Faculty of Engineering, Anadolu University, can be considered quite young.

TABLE 2

Attendance of Teachers at Language Testing Courses

Teachers' attendances at a course in language testing	Number of Teachers
Attended	2
Not attended	7

As it is seen in Table 2, with relation to the knowledge of language testing, 7 teachers indicated that they hadn't taken a course in this subject in college. There are only two teachers indicated that they had taken a course in language testing.

About this issue, Paterno (1960) points out that every teacher at one time or another has to prepare tests to measure the achievement of his/her pupils and the success of his/her teaching. She also adds that although a great deal of literature is available on the methodology of foreign language teaching and more and more suitable teaching material is being prepared for foreign learners, very little has been written or published on the preparation of English tests for use in foreign learners' classes. So, this means that we have been helped very little on the matter of testing achievement by the language testing courses in the teaching of a foreign language.

TABLE 3

Attendance of Teachers at in-service Courses or
Seminars Concerning Language Testing

Teachers' attendances at in-service courses or seminars in language testing	Number of Teachers
Attended	4
Not attended	5

As it is learned from Table 3, most of the teachers have never attended any in-service course or seminar concerning language testing. There are four teachers who once have attended an in-service course or seminar in language testing.

Present findings from Tables 2 and 3 seem to indicate that since the success of any formal language program is crucially linked with its testing philosophy and practice, the further training of the teachers in the theory of ELT and testing is of foremost importance.

TABLE 4

Opinions of the Teachers About the Present Tests

Comments of teachers on the appropriateness of present language testing situation	Number of Teachers
Good	3
Satisfactory	1
Fair	5

According to the Table 4, 5 teachers view the testing system as fair. On the other hand, 3 teachers think that the testing system is good. One teacher believes that the present language testing system is satisfactory.

It is certain that as having close and regular contact with students during the learning process,

the teachers will reveal the actual characteristics of the tests. Hints from the classroom teachers will significantly shape the various factors involved in the testing system. Consequently, it can be concluded from the fourth table that in their approach to present language testing situation, the majority of teachers did not give the impression that they were proud of their language tests.

According to Table 4, most of the teachers unanimously agree that multiple-choice, true-false, fill-in the blank, matching and short composition tests can be more useful in evaluation of student's success. This indicates that, except the short composition, the preferred types of tests are all classified in the objective marking category.

On the other hand, the majority of teachers view short-answer, transformation, addition and rearrangement tests as inappropriate. This shows that the semi-objective test types are not preferred by the most of the teachers. As a natural result of this tendency, teachers who view the essays as less useful type of tests are in the majority.

In the light of the findings from Table 5, it is worth noting that although the majority of teachers view objective tests as more useful, a good classroom test will probably contain both subjective and objective

TABLE 5

Opinions of the Teachers About the Present Tests

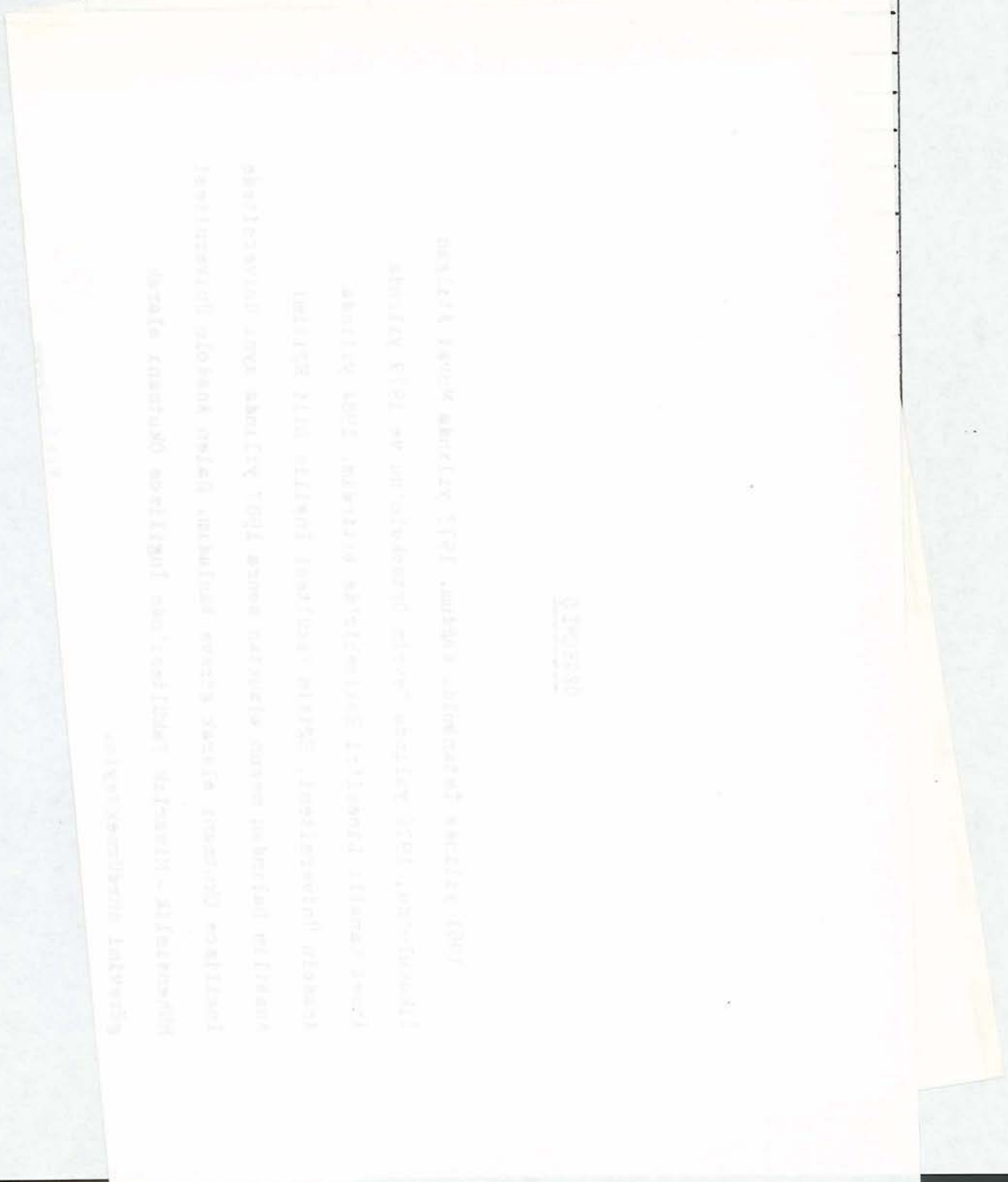
Test Type	Comment	Number of Teachers
Multiple-Choice	Appropriate	9
	Inappropriate	0
Fill in the blank	Appropriate	7
	Inappropriate	2
Short answer	Appropriate	4
	Inappropriate	5
Transformation	Appropriate	3
	Inappropriate	6
Addition	Appropriate	3
	Inappropriate	6
Rearrangement	Appropriate	4
	Inappropriate	5
True-False	Appropriate	7
	Inappropriate	2
Matching	Appropriate	8
	Inappropriate	1
Short Composition	Appropriate	5
	Inappropriate	4
Essay	Appropriate	1
	Inappropriate	8

test items, Heaton (1975:13). According to him the objective tests can never test ability to communicate in the target language nor can they evaluate actual performance.

TABLE 6

The Administering Frequency and Opinions of Teachers
on Certain Types of Language Tests

Preferences of Teachers	Administering Frequency	Number of Teachers
Multiple-Choice	Always Generally	



It can be concluded from the sixth table that although the preferred types of test items are mainly objective items, there are some inconsistencies in administering frequency of them. Because most of the teachers indicate that fill-in the blank, true-false and matching type of test items are never used. But as it is seen in Table 5, most of the teachers stated that those items were appropriate in evaluation of student's success. It is also interesting to note that although the most of the teachers agreed on the appropriateness of short composition in the previous table, a majority of teachers believe that they never use short composition according to the Table 6. All the teachers believe that the most frequently used tests are multiple-choice item types. It will also be seen in Table 6 that the semi-objective items are less frequently used and essays are never used types of tests.

Concerned with the Table 7, most of the teachers indicate that they never use language tests to measure students' success. On the same item, according to 4 teachers they always use language tests to measure students' success. Similarly, the number of teachers who never use language tests to diagnose students' learning difficulties is 5. On the same item, one teacher marks the option 'rarely' and 3 teachers mark the option 'sometimes'. As for the purpose of encouraging

TABLE 7

Opinions of the Teachers About the Purposes of
Language Tests

Purposes of Language Tests	Frequency of Administering Purposes	Number of Teachers
To measure students' success	Always	4
	Never	5
To diagnose students' learning difficulties	Sometimes	3
	Rarely	1
	Never	5
To encourage students for much and better learning	Generally	2
	Sometimes	1
	Rarely	1
	Never	5
To improve teaching methods	Sometimes	2
	Rarely	2
	Never	5
Only to assign the students some grade	Always	2
	Generally	2
	Sometimes	2
	Rarely	2
	Never	1

students for much and better learning, most of the teachers believe that they never use language tests for this purpose. Similarly, most of the teachers believe that language tests are never used for the purpose of improving teaching methods.

As might be expected only one teacher believes that language tests are never used only to assign the students some grade. Although there isn't a noticeable difference between the teachers who mark the other options, it can be said that a majority of teachers use language tests only to assign the students some grade. About the same issue, Oller(1987: 44) states that at a certain primitive level, tests in general, including language tests, are merely a means to the intermediate goal of assigning or justifying a grade. He also adds that this primitive view might be summed up in the phrase "guilt be association".

TABLE 8

Opinions of the Teachers About the Extent of Cooperation
Among Teachers

Frequency of consultations among classroom teachers for eliciting comments on the administered tests	Number of Teachers
Always	2
Usually	1
Sometimes	5
Seldom	1

According to the Table 8, it can be observed that most of the teachers consult other teachers for eliciting their comments. As it is seen in the table, only one teacher seldom does this consultation.

TABLE 9

The Extent of Cooperation Between the Teachers in
Constructing the Tests

Comments of the teachers on the necessity of a live cooperation	Number of Teachers
Necessary	8
Not necessary	1

As can be observed in Table 9, a majority of teachers believe the necessity of a live cooperation between the teachers. In other words, it can be said that as might be expected, most of the teachers unanimously view a live cooperation between the teachers as a vital necessity. However, one teacher doesn't consider this cooperation as a necessity.

On the other hand, it is clear, that although the teachers believe the necessity of a live cooperation, they do not always consult other teachers for eliciting their comments on the administered tests as it is seen in Table 8. About this issue Lowe (1975: 323) points out that quite often there is not enough coordination among English teachers in the same school, especially when a system of team-teaching is used. So, the lack of coordination produces serious inconsistency.

According to the Table 10, the most of the teachers believe that the calculation of the mean

TABLE 10

Opinions of the Teachers About the Calculation
of the Mean of the Test Scores

Comments of the teachers about the usefulness of the Mean of the test scores	Number of Teachers
Helpful	9
Not helpful	0

of the scores is helpful. This also indicates that all teachers view the calculation of the mean of the scores as the most efficient measure of central tendency.

TABLE 11

Opinions of the Teachers About the Interpretation of the
Test Results in Constructing the Forthcoming Tests

Comments of the teachers about the interpretation of the test results	Number of Teachers
Teachers who take the test results into consideration in construction of forthcoming tests	7
Teachers who does not take the test results into consideration in construction of forthcoming tests	2

As it is seen in Table 11, a majority of teachers indicate that they take the test results into consideration in the construction of forthcoming tests. This means

that according to the most of the teachers if test results are not dealt with properly, tests would not provide a valid basis for discrimination purposes. However, two teachers indicate that they do not take the tests results into consideration in the construction of forthcoming tests. It is certain that the teacher who wishes to know how well a test has worked, or how the scores it has produced are to be interpreted, needs to know how to deal with some simple statistics. According to Harrison (1983: 118) if a teacher is setting an achievement test that is to be used for important qualification purposes, he will take more time and care over the preparation of it, probably in collaboration with other teachers.

TABLE 12

Opinions of the Teachers About the Use of Test Results

Teachers' purposes for the use of test results	Frequency of Purposes	Number of Teachers
To find out students' learning difficulties	Always	6
	Generally	1
To get acquainted with students to orient them properly	Sometimes	7
To stimulate students for much and better learning	Generally	5
	Sometimes	1
	Rarely	1
To improve teaching methods	Generally	1
	Sometimes	1
	Rarely	1
	Never	4

Closely related to the eleventh table, in Table 12, the most of the teachers believe that they always use the test results to find out students' learning difficulties. On another item, a majority of teachers agree that these results are sometimes used to orient the students. According to five teachers, the test results are generally used to stimulate students for better learning. However, in the last item, just a reversal of these comments is noticeable. There are only two teachers who generally or sometimes use the test results to improve the teaching methods. But one teacher rarely and four teachers never use the test results to improve teaching methods.

About this issue Finocchiaro and Sako (1983: 36) point out that in order to determine the effectiveness of innovative teaching procedures grades obtained by students are often used either as one way of judging the efficacy of newer teaching techniques, or for recommending the introduction or suppression of teaching variables.

According to Table 13, 5 teachers do not consider their tests congruent with the aims and practices of the language teaching. However, 4 teachers believe that their tests are congruent with the aims and practices of language teaching. About this issue Heaton (1975: 1) emphasizes that both testing and teaching are so closely

TABLE 13

Opinions of the Teachers About the Appropriateness
of Language Tests with the Aims and Practices
of Language Teaching

Comments of teachers for their tests in terms of the aims and practices of the language teaching	Number of Teachers
Teachers who believe their tests are congruent with the aims and practices of the language teaching	4
Teachers who believe their tests are not congruent with the aims and practices of the language teaching	5

interrelated that it is virtually impossible to work in either field without being constantly concerned with the other. On the same parallel, Finocchiaro and Sako (1983:41) state that teaching and testing are the two sides of the same coin.

TABLE 14

Opinions of the Teachers About the Relationship Between
the Language Being Tested and the Way It Is Taught

Comments of the teachers about the language teaching issue related with the tests	Number of Teachers
Teachers who believe the language is being tested in the way it is taught	5
Teachers who believe the language is not being tested in the way it is taught	4

As it is seen in Table 14, most of the teachers think that the language is being tested in the way it is taught. On the other hand, four teachers believe that the language is not being tested in the way it is taught. It can be concluded from the present findings in Table 14, that although a majority of teachers assume their tests reflect actual teaching situations some of them don't believe this.

TABLE 15

Opinions of the Teachers About the Appropriateness of Their Language Tests to the Course Objectives

Comments of the teachers about the appropriateness of their tests to the course objectives	Number of Teachers
Appropriate	3
Inappropriate	6

As can be learned from Table 15, a majority of teachers view the tests inappropriate to the course objectives. There are only three teachers who think that their tests are appropriate to the course objectives.

The aim of the Department of English in the Faculty of Engineering is to teach the students enough English in a four-year program to enable them to follow the academic studies in their own fields. The order of priority of language skills can be listed as the

development of reading, speaking, writing and listening. That's why, it can be said that the students' cognitive skills are given the primary emphasis at this faculty. This tendency results in the passivization of their productive skills. Keeping this in mind, the teachers undoubtedly accept the fact that the tests do not correspond to the course objectives.

TABLE 16

Opinions of the Teachers About Some Aspects of Language Skills that the Tests Make Demands on

Comments of the teachers about the confines of language tests they prepare	Number of Teachers
Language tests are confined themselves to recognition	8
Language tests can measure more creative aspects of language	1

According to Table 16, most of the teachers believe that the tests confine themselves to recognition rather than more creative aspects of language. This means that the tests make demands on the cognitive skills.

According to Paterno (1960: 378) language learning is a complex thing and made up of many different aspects. It is a mistake to think that by measuring one aspect you measure all others, that any one aspect will give

an index to the whole. Therefore, it goes without saying that all the aspects of a language program to be measured.

TABLE 17

Opinions of the Teachers About the Way in which
Learner is Taken into Account on Tests

Comments of the teachers about evaluation of students' own performance on the tests	Number of Teacher
Teachers who believe the learner can evaluate his own performance	7
Teachers who believe the learner cannot evaluate his own performance	2

As can be observed in Table 17, the most of the teachers indicate that the learner can evaluate his own performance on the tests. This shows that according to a majority of teachers their tests allow formative evaluation by the students.

TABLE 18

Opinions of the Teachers About the Appropriateness
of the Tests in Terms of the Testees' Knowledge,
Expectations and Skills

Comments of teachers about the appropriateness of language tests in terms of students' knowledge, expectations and skills	Number of Teachers
Appropriate	7
Inappropriate	2

According to Table 18, a majority of teachers described the tests they prepare as appropriate for the proposed testees in terms of their knowledge, expectations and skills. Although two teachers indicate that their tests are inappropriate, most of the teachers agree that the tests meet students' expectations related with their capacities.

TABLE 19

Opinions of the Teachers About Their Tests'
Instructions and Tasks

Comments of the teachers on the clearness of their test instructions and tasks	Number of Teachers
Clear	9
Not Clear	0

As it is seen in Table 19, all the teachers believe that the instructions and tasks of tests are clear. About this issue Henning (1982: 36) points out that instructions of the tests must be clear, both to the students and to any test administrators using the test. On the same parallel, all teachers find nothing wrong with the test instructions.

As can be observed in Table 20, a majority of teachers indicate that the questions are ranked in terms of difficulty level as much as possible.

TABLE 20

Opinions of the Teachers About the Order
of Test Questions

Comments of the teachers about the test questions' progresiveness in difficulty	Number of Teachers
Progressive in difficulty	8
Not progressive in difficulty	1

Only one teacher believes that the test questions are not progressive in difficulty.

It can be concluded from this table that although there is one teacher who disagrees with the issue of hieararchy in terms of difficulty, most of the teachers have a positive view on this topic. According to Grittner (1969: 342) in presenting test items it is considered good practice to begin with the simplest items and proceed to the more difficult items. Similarly, Finocchiaro and Sako (1983: 211) also state that the arrangement of items should be made according to their difficulty level.

According to Table 21, most of the teachers believe that tests avoid items which provide more than one acceptable answer. On the same item, only one teacher believes that the questions provide more than one acceptable answer. Although a majority of teachers believe that their questions do not provide

TABLE 21

Opinions of the Teachers About the Avoidance of Questions
that Provide More Than One Acceptable Answer

Comments of the teachers about the existence of questions with more than one acceptable answer	Number of Teachers
Questions provide more than one acceptable answer	1
Questions do not provide more than one acceptable answer	6
Questions generally avoid providing more than one acceptable answer	2

more than one acceptable answer it is not easy to prepare concrete tests as measurement tools. Because some uncontrolled factors may cause discrepant results. According to Oller (1979: 242) in writing multiple-choice items it is essential that the correct answer or the best choice must fit the facts well and further it must be better than the other alternatives offered. That's why, the facts ought to be clear on the basis of the text, or they should not be used as content for test items.

As can be seen in Table 22, the most of the teachers indicate that the tests are not being appropriately pre-tested. This indicates that the most of the shortcomings of tests are caused by the lack of pretesting application. Because the pre-testing process can decide the preciseness

TABLE 22

Opinions of the Teachers About pre-testing
of Language Tests

Comments of teachers about pre-testing of their language tests	Number of Teachers
Appropriately pre-tested	1
Not pre-tested	8

of implications of the tests. About this issue, Oller (1979: 8) points out that, especially in constructing the multiple-choice tests the items need to be pretested, preferably on some group other than the population which will ultimately be tested. Then, the items need to be statistically analyzed so that non-functional or weak items can be revised before they are used and interpreted in ways that affect learners.

TABLE 23

Opinions of the Teachers About General Test Organization

Comments of teachers about the adequacy of general test organization	Number of Teachers
Adequate	5
Inadequate	4

According to Table 23, the most of the teachers view general test organization as adequate. However,

4 teachers find the general test organization inadequate. So, it can be concluded that the present testing approach makes provision for adequate test organization. For obvious reasons it is worth noting that although the majority of teachers did not give the impression that they were proud of their language tests, they find the general test organization adequate.

The overall conclusions drawn from the answers given by the teachers, it is evident that there are some points where the subjects do not agree with each other. It goes without saying that the teacher has the ultimate responsibility for deciding what he will test and how he will do it, or indeed, whether to test at all. He is influenced by the aims and needs of the students he is teaching, the course book he is using, the demands of the school and the system and so on, and must therefore devise tests to fit these conditions. It may be said that if a person writes a test by himself and puts it aside for a few days and comes back to revise it later, he will certainly find all kinds of errors and inconsistencies which he had not noticed as he worked the test out in detail. But this is no substitute for the comments of an interested colleague, who will see the test material from a different viewpoint and will point out ambiguities and possibilities for error which the test writer cannot see: he is too committed to his idea and has

too detailed a view of it by the time he has worked it out. That's why, in the following section the analysis will be contrasted by some observations of the sample test.

4.2. Some Observations on the Sample Test

In this section, a midterm test prepared by the all English language teachers of the Faculty of Engineering, and administered in the 1989-90 academic year will be analyzed (c.f. Appendix B-1).

This test was administered to students soon afterwards the questionnaire given to teachers. Since the test is designed to provide an objective testing basis, there aren't any subjective parts in this test. Therefore, this test can be described as an objective achievement test.

Discussion of our testing mistakes will be based on two checklists of Henning (1982: 33) and Grittner (1969: 342), (c.f. Pages 33-36). According to Henning (1982) if the following mistakes of testing are avoided, an improved testing program can be achieved. That's why, to enable the exhibition of the current testing situation, by an analysis of a midterm test, the following table (Table 24) was constructed with the help of checklist.

In the Table 24, the testing mistakes are grouped in three categories as general examination characteristics, item characteristics and test validity concerns. These categories are evaluated by some numbers that are used to represent the appropriateness of the test from the viewpoint of avoided testing mistakes. The appropriateness of the midterm test from the point of avoided testing mistakes is enumerated as follows: 1. Poor, 2. Fair, 3. Satisfactory, 4. Good, 5. Excellent.

As Henning (1982) points out the other problems of testing may also be cited but the ones given here are certainly the more common. Therefore, the following table can also be used as a guide for EFL teachers in the preparation of their own examinations.

As it is seen in the Table 24, the midterm test administered at the Faculty of Engineering in 1989-90 academic year can be considered as poor from the point of avoidance of the items that are too easy or too difficult. Because it is clearly observed in Table 25 that the most of the items have an index of difficulty (or a facility value) higher than .6 or 60%. About this issue Heaton (1975: 173) emphasizes that an average facility value of .5 or 50% may be desirable for many achievement tests.

TABLE 24

The Appropriateness of the Midterm Test Administered
at the Faculty of Engineering in 1989-90 Academic
Year From the Point of Avoided Testing Mistakes

Characteristics of the test	The common mistakes of testing to be avoided	1	2	3	4	5
GENERAL EXAMINATION CHARACTERISTICS	Test which is too difficult or too easy					
	An insufficient number of items					
	Redundancy of test type					
	Lack of confidence measures					
	Negative washback through non-occurrent forms					
ITEM CHARACTERISTICS	Trick questions					
	Redundant wording					
	Divergence cues					
TEST VALIDITY CONCERNS	Mixed content					
	Wrong medium					
	Common knowledge					

1. Poor, 2. Fair, 3. Satisfactory, 4. Good, 5. Excellent

TABLE 25

Facility Value of the Items of Midterm Test

Item Number	Facility Value	
	Group I	Group II
1	.67	.70
2	.97	.76
3	.75	.80
4	.37	.48
5	.24	.80
6	.29	.76
7	.97	.76
8	.78	.56
9	.60	.94
10	.19	.87
11	.60	.25
12	.50	.74
13	.54	.66
14	.48	.30
15	.48	.66
16	.50	.69
17	.35	.84
18	.54	.89
19	.24	1.00
20	.05	.66
21	.86	.89
22	.80	.40
23	.40	.33
24	.56	.46
25	.60	.43
26	.60	.76
27	.40	.60
28	.75	.80
29	.67	.50
30	.67	.80
31	.50	.25
32	.78	.70
33	.70	.74
34	.94	.90
35	.40	.70
36	.78	.38
37	.43	.58
38	.43	.87
39	.30	.90
40	.40	.87

According to the Table 25, it is obvious that most of the items are fairly easy since a majority of students taking the test answered them correctly. On the other hand, the facility value of 8 items in first group and 5 items in the second group are below .4 or 40%. This indicates that the facility value of the items in both group can be considered as out of the average range of difficulty levels. About the same issue Harrison (1983: 128) indicates that the usual aim of test writers is to achieve fairly even and middling facility indices right through the test, ranging from about .40 to about .60. Therefore, it can be concluded that in the midterm test there is an accumulation of scores at the higher ends of the scoring range. As a result of such effect, there is information loss and reduced capacity of the test to discriminate among students in their ability. That's why, the midterm test can be considered as poor from this point as it is introduced in Table 24.

On the other hand, it is obviously seen that the sample test (c.f. Appendix B-1) includes forty items. According to Henning (1982: 33) test reliability is directly related to the number of items occurring on the test. He also adds that it is always difficult to achieve acceptable reliability with less than 50 items. About this issue, Finocchiaro and Sako (1983: 238)

point out that as a general rule, increasing the length of a test improves its reliability so that a 120-item test generally is more reliable than a 60-item test. This has been confirmed empirically as well as statistically. Therefore, as it is seen in Table 24, this is considered as a common mistake of testing to be avoided. And for obvious reasons our sample test can be accepted as a short one and thus poor from the point of avoided testing mistakes.

As it is seen the midterm test was constructed in three parts (See Appendix B-1). The first part, Part A, includes 15 items about grammar and usage. Part B includes three reading passages and 10 comprehension questions. Last part is Part C, and it was designed to test vocabulary with 15 items. As Henning (1982) points out in testing general language proficiency, it is common practice to devise a battery of subtests to ensure that all important language skills are covered by the test as a whole. He also adds that inclusion of subtests of grammar accuracy, vocabulary recognition, and composition writing "leaves no room" for a subtest of listening comprehension. But it is clear that our sample test is an achievement test. For this reason and because of the characteristics of achievement tests, normally tests can involve simple recall of prelearned material and rearrangement of

prelearned material. The text book used at the Faculty of Engineering is based on the functional-notional approach. According to Finocchiaro and Sako (1983: 264) the functional-national orientation has already had considerable influence on the ways in which communicative performance in a language described, the types of classroom exercises and other learning experiences that build toward this objective, and the techniques used to measure its attainment. That's why, the sample test can be criticized from the point that it doesn't have all the important language skills as a whole. And since it measures some language components and some skills, in Table 24, it can be considered as fair in redundancy of test type from the point of avoided testing mistakes.

As it is seen the sample test was constructed in multiple-choice item type (See Appendix B-1). That's why the scoring procedure of this test is considered as objective type. And in the scoring of the test computers were used (See Appendix B-2). Consequently, the data collected through the computers can be used to estimate the level of confidence that we may place in the test result when it is applied to various situations. But the results of the questionnaire showed that (c.f. Table 12) the test results are not interpreted appropriately. Because

some teachers indicate that they never use the test results to improve teaching methods, sometimes or rarely use to stimulate students for much and better learning and sometimes use to orient students properly. This indicates that although locally developed tests are used for important evaluative decisions and estimates of reliability and validity should be provided, most of the teachers do not always take these characteristics into consideration and find the data sheets satisfactory for the further use of the tests. Therefore, in Table 24, the avoidance of the lack of confidence measures can be evaluated as poor in the analysis of the sample test.

When it comes to the point of avoidance of negative washback through non-occurent forms, the sample test generally achieves this by using appropriate structures of the language. About this issue, the option (b) in the item number (13) can be revised (See Appendix B-1). According to Henning (1982) while it is necessary that options include incorrect forms as distractors, it is best if these forms have some possible appropriate environment in the language. That's why, the sample test can be considered as satisfactory in avoiding non-occurent forms in the Table 24.

As it is seen in Table 24, the sample test can be considered as good in avoiding trick questions.

In the sample test, the stem of the item number(3) can be revised (See Appendix B-1). Because in the dialogue, the passenger uses the verb 'Lose' but the customs officer changes it to the verb 'Forget'. Although this cannot be considered as a trick question, this kind of errors impair the quality of a test.

On the other hand, it can be said that the sample test generally avoids redundant wording. Only the item (33) can be revised by putting the word 'Appointment' in the stem (See Appendix B-1). Therefore, as it is seen in the Table 24, the sample test is evaluated as good because there is no any other needless repetition. This shows that, in constructing the test all the teachers are aware of the fact that items with redundant wording greatly reduce the efficiency of a test.

It may be said that there is no option which gives greater length or specificity of information in the sample test. This indicates that the sample test does not provide some cues regarding the choice of the correct option. So, it can be concluded that option divergence was successfully avoided in constructing the sample test. Therefore, in the Table 24, it is evaluated as excellent in avoiding divergence cues.

It is certain that sometimes tests have been claimed to measure something different from what many

of their items are actually measuring. Similarly, in the sample test the item, number (6), was placed in the group of questions about grammar and usage (See Appendix B-1). But the options indicate that they actually measure vocabulary recognition. According to Henning (1982) these kind of inconsistencies make for invalid tests. Since the other items of the sample test generally avoid mixed content, it can be considered as good in Table 24.

As it is seen in the sample test the reading-comprehension questions are all formed in multiple-choice item type (See Appendix B-1). About this issue Henning (1982) points out that sometimes reading-comprehension questions require accurate written responses to show comprehension of the passage. He also adds that research has indicated that such tests are invalid in the sense that they measure something other than what they are intended to measure. That's why, it can be concluded that in the sample test the response medium is representative of the skill being tested. So, it can be shown in Table 24 that the sample test perfectly avoids wrong medium.

Similarly, items that require common-knowledge responses are also avoided in the sample test. Therefore, the sample test is considered as excellent in avoiding common knowledge from the point of avoided testing mistakes in the Table 24.

In sum, as it is seen in the Table 24, the midterm test can generally be criticized as poor in terms of the general examination characteristics. However, when it comes to the item characteristics and test validity concerns, the sample test can be considered as good. Of course, the problems stated in the Table 24 are not the all problems of English language testing. That's why, the following checklist of Grittner (1969: 342) can also be helpful in exhibiting the present situation of the sample test. Because his checklist is particularly related to the construction of achievement tests.

As it is seen in Table 26, the shaded areas represent the degree of appropriateness of the sample test to the basic principles of achievement testing. The checklist is divided by some numbers as follows: 1. Poor, 2. Fair, 3. Satisfactory, 4. Good, 5. Excellent.

According to Grittner (1969) normally achievement tests can involve simple recall of prelearned material. The sample test can be considered as satisfactory in terms of providing simple recall of prelearned material. Because the sample test was constructed to determine whether the units 6 and 7 of the textbook "Exchanges" have been mastered by the second year students. That's why the sample test utilize the vocabulary and the grammatical structures of the text book. But when it comes to listening and speaking skills, the

TABLE 26

The Appropriateness of the Sample Test to the
 Checklist of Basic Principles for Constructing
 Achievement Tests (Grittner 1969: 342)

Basic Principles of Achievement Testing	Appropriateness of the Sample Test to the Principles of Achievement Testing				
	1	2	3	4	5
Simple Recall of Prelearned Material					
Rearrangement of Prelearned Material					
Beginning with the Simplest Items					
Proceeding to the more difficult Items					
Selecting Problem sounds, Structures and Word-Order Sequences					
Avoiding Tongue Twisters and Absurdities					
Including at least one distractor, two plausible but incorrect items and one unambiguous item which correctly answers the question					
Directions on the Test should be in Native Language but the questions should be in the target language					

1. Poor 2. Fair 3. Satisfactory 4. Good 5. Excellent

sample test fails in measuring them. As it is mentioned earlier (c.f. Table 24) the textbook used at the Faculty of Engineering is based on the functional-notional approach. So, there are a lot of activities in which communicative performance takes part. But, it can be said that the sample test does not utilize all language skills as a whole. Similarly, most of the teachers indicated that the student's cognitive skills are given primary importance at the Faculty of Engineering (c.f. Table 15). Consequently, this tendency results in passivization of the students' productive skills. For obvious reasons the sample test is considered as satisfactory in Table 26.

Grittner (1969) also indicates that one of the basic principles for constructing achievement tests is rearrangement of prelearned material. As can be observed in the sample test, questions about the vocabulary recognition usually stay in the vocabulary range of text book material (See Appendix B-1). To be creative, synonyms and antonyms of some words can be given. By doing this, we can help the students increase their vocabulary range. And also some pictures, short answer or completion type of items can be used. The principle "Do not ask questions beyond syllabus" does not mean that we should ask the same text book material. Our failure stems from that misconception. The result is that when the students see an unfamiliar

word among the choices they eliminate it right away, because that word is not in the printed material of the text book. About this issue, most of the teachers indicated in the results of the questionnaire that the tests they prepare confine themselves to recognition rather than more creative aspects of language (c.f. Table 17).

It is certain that if we do not manipulate the text book material we give a test measuring knowledge only. Because manipulation leads us application, that is to be productive. For example, in the sample test, item 12 is almost the same as the sentence in the text book (See Appendix B-1). That's why it can be said that the students have become test minded and they have confined themselves to the text book material. For these reasons, the sample test can be considered as fair in Table 26, in terms of the rearrangement of prelearned material.

According to Grittner (1969) in presenting test items it is considered good practice to begin with the simplest items and proceed to the more difficult items. But it is evident in our sample test that the items were not arranged in order of difficulty with the easiest items first (See Appendix B-1). About this issue, it is interesting to note that most of the teachers indicated in the questionnaire that their test questions are progressive in difficulty (c.f. Table 20).

There is only one teacher who stated that their test questions are not progressive in difficulty. This indicates that there is an inconsistency between the opinions of the teachers about the order of test questions and the present application of a test. It is certain that placing difficult items early in the test may cause students to spend much time on these and prevent them from reaching items they could easily answer. Thus it can be said that improper arrangement may influence validity by having a detrimental effect on student motivation. Therefore it can be concluded that in Table 26 the principles beginning with the simplest items and proceeding to the more difficult items can be considered as poor in the analysis of the sample test.

As it is seen in the sample test there is no oral or listening comprehension subtests (See Appendix B-1). That's why, the principles given in the checklist as selecting problem sound and avoiding tongue twisters cannot be taken into consideration in the analysis of the sample test. As it is seen in the sample test it can be said that selecting problem structures and word-order sequences is satisfactorily achieved on the same parallel with the content of text book. In terms of the avoiding absurdities, it is clearly seen that the third question of the sample test gives two different verbs and forms a kind of ambiguous

statement. That's why, it is shown as fair in the Table 26.

As it is seen in the sample test the most of the distractors are easily eliminated by a majority of students (c.f. Table 25). This indicates that the distractors were not plausible enough. It is certain that to the students who has not achieved the learning outcome being tested, the distractors should be at least as attractive as the correct answer preferably more. If a distractor is not selected by anyone, it should be eliminated or revised. For example in the second test group the all distractors of item 19 were eliminated by the whole group taking the test. Similarly, the most of the items have a facility value of 60% or higher. Therefore, the sample test is considered as poor in setting up most plausible and educationally sound distractors, in Table 26.

Present findings from the answers given by the teachers to the questionnaire seem to indicate that all teachers find nothing wrong with their tests' instructions. Similarly, as it is seen in the sample test, the directions on the test are in the native language and the questions are in the target language. This means that the sample test can be considered as good in Table 26, in terms of basic principles of achievement testing related to the test directions.

But it can be worth noting that there is no any page notice like "turn over" or "continue to the next page" on the text booklets of the sample test (See Appendix B-1). Sometimes, in the previous tests, some students missed the items on the back pages. Therefore, it can be said that notices mentioned above can be useful in preparing better directions for the tests.

CHAPTER V

DISCUSSION AND SUGGESTIONS

5.1. Discussion

The main purpose of this study was to elicit information about the most outstanding shortcomings of the existing testing system and to establish some suggestions for more feasible ways of increasing facility in language testing situation at the Faculty of Engineering.

In the light of comments directed by the teachers and with the present findings derived from the sample test, some of those shortcomings can be discussed as follows.

It can easily be observed that there is a need for the further training of the teachers in English

language testing at the Faculty of Engineering (c.f. Table 2 and 3).

It is generally acknowledged that since the success of any formal language program is crucially linked with its testing philosophy and practice, teachers as the central body of language teaching process, have an obligation to improve the efficiency of existing tests by making contributions.

Purposes of language tests are not determined appropriately (c.f. Table 7).

The results of the questionnaire showed that most of the teachers never use language tests for the goal of improved learning and instruction. According to Oller (1987: 44) if the language tests are used with only the purpose of assigning students some grade this is a primitive view.

Test results are not interpreted appropriately (c.f. Table 12).

Test scores by themselves have little meaning. As Finocchiaro and Sako (1983: 258) point out the test scores interpreted and used; a) to find out students' learning difficulties, b) to get acquainted with them for orientation, c) to stimulate students for better learning, d) to improve teaching methods, e) to develop course curriculum.

The tests are not appropriate to the course objectives (c.f. Table 15).

The aim of Department of English in the Faculty of Engineering is to teach the students enough English in a four year program to enable them to follow the academic studies in their own fields. The order of priority of language skills is reading, speaking, writing and listening.

The tests are not congruent with the aims and practices of language teaching (c.f. Table 13).

Language teaching does not mean to provide students with a mere knowledge of vocabulary and grammar. Students should be furnished with all elements and skills of the language.

The tests confine themselves to recognition rather than more creative aspects of language (c.f. Table 16).

A good language test may contain either recognition-type test items or production-type items, or a combination of both, since each type has its unique functions.

The tests are not appropriately pre-tested (c.f. Table 22).

It is necessary to make predictions about the preciseness of the test. Therefore, pre-testing is useful in selecting appropriate item types.

In the light of evidences provided from the questionnaire and the sample test, it is clear that there are some points where the subjects do not agree with each other. Although most of the teachers believe that their test questions are progressive in difficulty (c.f. Table 20), it is observed in the sample test that the questions were not arranged in order of difficulty with the easiest items first (c.f. 4.2.). Consequently, this may influence the validity of the tests.

In sum, It is obvious that without a sound evaluation we cannot make decisions about the instructional objectives, we cannot be sure what we should teach and to what extent, where our programs failed and what the learning difficulties are. The purpose of using tests and other evaluation instrument during the instructional process is to guide and direct student learning and monitor progress toward course objectives. Therefore, the teachers have great responsibilities to eliminate at least some of the constraints presented above for a better testing situation at the Faculty of Engineering. Certainly, it is a tiring process. But the outcomes will lead to more desirable tests. Because measurement and evaluation reflect the quality and the direction of education.

5.2. Suggestions for further studies

Although this study was carried out by use of a questionnaire and considering the requirements for the research, the following points will be beneficial for further studies. We can tabulate suggestions in the following form.

1. This research only covered the teachers of English at the Faculty of Engineering. A similar study can be conducted for the other faculties where teaching of English as a foreign language has been carried out.
2. The questionnaire in this study can be further developed and new techniques can be used to obtain information about the present testing situations.
3. Further test types can be incorporated in the analysis of achievement tests so that the boundaries of research can be enlarged.
4. A research can be attained by an analysis of the items in the previous tests to bring out some criteria on various aspects of the current tests from the point of view of test construction.
5. In the light of information given in this study, some model tests can be developed

and applied to different groups of students to find out the variations with the accustomed systems of language testing.

REFERENCES

- ALDERSON, J.C. "Testing ESP in a Communicative Framework",
Course Handout (1980), University of Lancaster.
- BROWN, H.D. (1987) Principles of Language Learning and Teaching, Prentice-Hall, Inc.
- CAROLL, J.B. "Testing the English Proficiency of Foreign Students", Teaching English as a Second Language (1961), Arlington: Center For Applied Linguistics.
- CLARK, J.L.D. (1972) Foreign Language Testing: Theory and Practice, Philadelphia: The Center For Curriculum Development, Inc.
- DAVIES, A. "The Construction of Language Tests", Testing and Experimental Methods (1978), London: Oxford University Press.

- FINOCCHIARO, M. and S. Sako. (1983) Foreign Language Testing: A Practical Approach, New York: Regents Publishing Co.
- GRITTNER, F.M. (1969) Teaching Foreign Languages. New York: Harper and Row Publishers.
- GRONLUND, N.E. (1976) Measurement and Evaluation in Teaching. New York: Collier Macmillan Editions.
- HARRIS, D.P. (1961) English Testing Guidebook. Washington, D.C.: The American University Language Center.
- HARRISON, A. (1983) A Language Testing Handbook. London: Macmillan Press Limited.
- HEATON, J.B. (1975) Writing English Language Tests. London: Longman.
- HENNING, G. "Twenty Common Mistakes for EFL Teachers to Avoid", FORUM (1982) 20/3: 33-37.
- KOHONEN, V. "Some Current Problems in Language Testing", Kielikeskusuutisia (1976), 4: 33-42.
- KRASHEN, S.D. (1982) Principles and Practice in Second Language Acquisition. Oxford: Pergamon Press.
- LADO, R. (1961) Language Testing: The Construction and Use of Foreign Language Tests. New York: Mc Graw-Hill.

- LOWE, J.E. "Are You Fair In Your Grading?", FORUM (1975), 13/3-4: 322-327.
- MADSEN, H.S. (1983) Techniques in Testing. Oxford: Oxford University Press.
- NOLL, V.H. and D.P. Scannell (1972) Introduction to Educational Measurement. Boston: Houghton Mifflin Company.
- OLLER, J.W.Jr. (1979) Language Tests At School. London: Longman Group Limited.
- . "Practical Ideas for Language Teachers from a Quarter Century of Language Testing", FORUM (1987), 25/10: 42-55.
- PATERNO, A. "Foreign Language Testing", The Manila Secondary Teachers' Symposium (1973), Boston: Allyn and Bacon, Inc.
- PILLINER, A.E.G. "Subjective and Objective Testing", Language Testing Symposium (1973), Boston: Allyn and Bacon, Inc.
- SPOLSKY, B. "Language Testing: The Problems of Validation", Readings on English as a Second Language (1972) K. Croft. (ed.). Cambridge, MA: Winthrop Publishers.

TUCKMAN, B.W. (1975) Measuring Educational Outcomes: Fundamentals of Testing. New York: Harcourt Brace Jovanovich, Inc.

VALETTE, R.M. (1977) Modern Language Testing. New York: Harcourt Brace Jovanovich, Inc.

APPENDIX A-1

The Questionnaire Used to Obtain Information About the
Extent of Agreement or Disagreement on the
Appropriateness of English Language Tests
Among Teachers at the Faculty of Engineering

Dear Colleague,

The purpose of this questionnaire is to evaluate various aspects of the present language testing system at The Faculty of Engineering, Anadolu University. I would appreciate your opinions on identifying the strengths or weaknesses of the language tests administered in our faculty and would therefore be grateful if you reply the questionnaire in full. Please, DON'T write your name. Because the interviewees will remain anonymous.

Thank you for your kind co-operation in completing this questionnaire. If you have further comments about the questions, please don't hesitate to write to me.

Instructor Ateş Dalyan

APPENDIX A-1 (CONTINUED)

1. How long have you been teaching English?
 - a. 1-4 years
 - b. 5-9 years
 - c. 10-14 years
 - d. 20-30 years

2. When in college, did you take any course concerning language testing?
 - a. Yes
 - b. No

3. Have you ever attended any in-service course or/and seminar concerning language testing?
 - a. Yes
 - b. No

4. What is your comment on the appropriateness of the present language testing situation at the Faculty of Engineering?
 - a. Excellent
 - b. Good
 - c. Satisfactory
 - d. Fair
 - e. Poor

5. What types of language tests, do you think can be more useful at the Faculty of Engineering?

APPENDIX A-1 (CONTINUED)

- | | | |
|----------------------|---------|--------|
| a. Multiple-Choice | () Yes | () No |
| b. Fill in the Blank | () Yes | () No |
| c. Short Answer | () Yes | () No |
| d. Transformation | () Yes | () No |
| e. Addition | () Yes | () No |
| f. Rearrangement | () Yes | () No |
| g. True-False | () Yes | () No |
| h. Matching | () Yes | () No |
| i. Short Composition | () Yes | () No |
| j. Essay | () Yes | () No |

6. What types of language tests do you use at the Faculty of Engineering, and how frequently?

F R E Q U E N C Y

- | | <u>Always</u> | <u>Generally</u> | <u>Sometimes</u> | <u>Rarely</u> | <u>Never</u> |
|----------------------|---------------|------------------|------------------|---------------|--------------|
| a. Multiple-Choice | () | () | () | () | () |
| b. Fill in the Blank | () | () | () | () | () |
| c. Short Answer | () | () | () | () | () |
| d. Transformation | () | () | () | () | () |
| e. Addition | () | () | () | () | () |
| f. Rearrangement | () | () | () | () | () |
| g. True-False | () | () | () | () | () |
| h. Matching | () | () | () | () | () |
| i. Short Composition | () | () | () | () | () |
| j. Essay | () | () | () | () | () |

APPENDIX A-1 (CONTINUED)

20. Do you think the questions in your language tests are progressive in difficulty insofar as possible?

a. Yes

b. No

21. Do you believe the language tests you prepare avoid confronting the students with questions that provide more than one acceptable answer?

a. Yes

b. No

c. Generally

22. Are your tests being appropriately pre-tested?

a. Yes

b. No

23. Do you find the general test organization adequate?

a. Yes

b. No

24. If you have any other comments please state here:

.....
.....

APPENDIX B-1: A Midterm Test Group B

ANADOLU ÜNİVERSİTESİ
 MİLLÎ EĞİTİM BAKANLIĞI
 BAĞIRCIYI DÖNEMİ İNGİLİZCE 1. ARA SINAVI

- A. Find the grammatically correct answer.
 (Gramer açısından doğru cevabı bulunuz.)
1. Student: I went to bed late yesterday night and I couldn't get up early, so I missed the exam. I wish I
- a. didn't go to bed late b. hadn't got up early
 c. had gone to bed early d. had got up late
2. Student: I didn't know what the vocabularies mean , so I didn't answer the questions in the exam last week.
 Teacher: If you the vocabularies, you answered the questions.
- a. had studied/would have b. studied/would have
 c. had studied/would d. studied/would
3. Passenger: Excuse me sir. I've lost my passport, can I get on the plane without passport?
 Customs Officer: I'm sorry sir, you can't. You forgot it.
- a. shouldn't b. should had
 c. should have d. shoudn't have
4. A: We are going to Bursa. Are you coming with us?
 B: I'm sorry, I'm not going with you because I have an exam tomorrow. I wish I with you.
- a. had gone b. was going
 c. am going d. can go
5. A: getting up early?
 B: It makes me feel bored.
- a. What does it make b. How do you feel about
 c. How does it make d. What makes you
6. A: Have you ever chopped onions?
 B: Yes. Many times. But, it makes me want to
- a. laugh b. bored
 c. cry d. amused
7. A: Fikret won the pools and he bought a big house in the country.
 B: I know that. But if I him, I buy a small house near the sea.
- a. were/will b. was/would have
 c. was/will d. were/would

APPENDIX B-1 (CONTINUED)

8. A: How did the exam go?
 B: The questions were too difficult, so I.....
 answer some of them.
 a. could b. would
 c. should d. shouldn't
9. A: Hello, Tom. Where are you going?
 B: I'm going to apply for a new job. In fact,
 I for jobs for a month.
 a. am applying b. applied
 c. have been applying d. was applying
10. We went to a travel agency, we were planning a holiday.
 a. although b. because
 c. but d. and
11. A: Hello. Where are you going?
 B: I'm going to the hairdresser to
 a. cut my hair b. cut me my hair
 c. have cut my hair d. have my hair cut
12. We booked the holiday we paid a deposit of 100.000TL.
 a. although b. because
 c. but d. and
13. A: She doesn't eat fish.
 B:
 a. Adil does too b. So Adil does
 c. Nor does Adil d. So does Adil
14. A: You look ill today. What happened to you?
 B: I'm not feeling well.
 A: You go out.
 a. had better not b. hadn't better not
 c. had better d. hadn't better
15. Customs Officier: Now, you should complete this form and
 write down.....things:your name.age.
 address and occupation.
 Passenger : O.K. Sir. I'll do it..
 a. these b. those
 c. this d. that

APPENDIX B-1 (CONTINUED)

Read the paragraph and choose the correct answer.
/ Ağagıdaki soruları parçaya göre cevaplayınız/



Many people have begun to save energy. Paul Moore rides his bicycle to work to save gas. He began eight months ago. He carries his briefcase in the basket on the bicycle and heads for the office. He gets there in 30 minutes.

Paul has lost weight and become more healthy. He sleeps better at night. He enjoys the fresh air and sunshine. In bad weather he prefers bus and his car to his bicycle to go to work.

VOCABULARY:

ride: binmek

briefcase: çanta

lead: ilerlemek, gitmek, yönelmek

basket: sepet

lose weight: kilo vermek

outdoors: açık hava

6. a. Paul wants to save gas | b. He leaves his briefcase home
c. It takes an hour to get to the office d. He rides too fast
7. He feels better now because -----
a. He goes to the office on a bicycle b. He drives too fast
c. He goes to the office on foot
d. He carries the basket on the bicycle
8. In bad weather -----
a. He prefers his bicycle b. He enjoys riding his bicycle
c. He likes to go by bicycle d. He goes by bus and his car

Marcus and Beth Wilkins have finished their vacation. They have spent two weeks in California. They have visited friends and relatives in many cities and towns. They have had a very good time.

Today, they have to go home. They have waited 40 minutes for a taxi. They have to get to the airport. Their plane leaves in 15 minutes.

VOCABULARY:

vacation: tatil

relative: akraba

19. a. The Wilkins are going for a vacation
b. The Wilkins have come to the end of their vacation
c. The Wilkins have started their vacation
d. The Wilkinsons will start their vacation
20. a. They haven't enjoyed their vacation
b. They have waited many hours for a bus
c. They have visited many places and people
d. They have to stay two weeks in California
21. a. They are on their way home
b. They have to go to the bus station
c. Their plane leaves in two hours
d. Their train hasn't arrived yet

APPENDIX B-1 (CONTINUED)



My name is June Taylor. My daughter Sara and I live in Fairbanks, Alaska. I am a working mother. I work as a camerawoman for a TV station in Fairbanks. My job is interesting, so I enjoy it. I didn't always work or plan to work, but I started to work ten years ago. This is my story. I married Charles Taylor in 1966 when we were very young. Charles was a salesman. He liked to move. That's why we moved to a new city every year. We lived in Chicago, Detroit and Miami. We lived far from our families. Charles wanted a large family and so did I. I didn't plan to get a job. I planned to stay home with children.

Sarah was born in Miami. Charles and I were very happy. I stayed home with Sara. We moved to Alaska the next year. I didn't have a job, but I had a hobby. I liked to take pictures of Sara and I learned to take very good pictures. I started to sell them, but it wasn't a real job. I always worked at home.

Charles suddenly became very ill in 1979. He had cancer and he died after six months. I had no money, no education and no friends. Then a TV station in Fairbanks saw my pictures. They asked me to work for them. I am very lucky to have the job, because I can earn enough money for Sara and me. We are independent and I don't worry about the future any more.

VOCABULARY:

salesman: pazarcı	cancer: kanser
move : taşınmak	die: ölmek
plan : planlamak	education: eğitim
take picture: fotoğraf çekmek	lucky : şanslı
suddenly: aniden	earn : kazanmak
become : olmak	independent: bağımsız
worry : üzülme	future: gelecek

22. Which sentence about June is false?
- She moved to Fairbanks to find a job.
 - She thinks she'll have a good future.
 - She enjoys her job because it is interesting.
 - Her husband died of cancer about ten years ago.
23. June and Charles moved every year because -----.
- he was a salesman and travelled.
 - they lived far from their families.
 - he wanted to move to different places.
 - they wanted a large family.
24. Before Charles became ill in 1979, June was -----.
- working hard at the TV station.
 - planning to work at the TV station.
 - taking some good pictures at home.
 - staying home with the children.
25. When June married Charles, she didn't plan to work. This means -----.
- she didn't want to get a job.
 - she didn't need to get a job.
 - she didn't try to get a job.
 - she wanted to be a housewife.

APPENDIX B-1 (CONTINUED)

C. Choose the following correct vocabulary for the blanks.
(Boşluklar için en uygun kelimeyi seçiniz.)

- B
26. Wife: oh. We have lots of dirty clothes.
Husband: We can have them washed by the washing machine.
In this conversation, husband..... to use the machine.
a. refused b. confused c. rejected d. offered
27. Customer: How much does that radio cost?
Assistant: It is 250.000TL.
Customer: It is too expensive for me. Would you please.....
the price to 200.000TL.
a. increase b. cut c. cheap d. borrow
28. A: I have a pain in my arm.
B: Why is that?
A: Because the nurse has just given me
a. an injection b. a leaflet
c. some medicine d. some advice
29. Ali: Have you seen the movie at Ari Cinema?
Ahmet: What's on?
Ali: LAMFADA.
Ahmet: O.K. Who is going to book the tickets?
Ali: Don't worry. I'll book the tickets.
In this dialogue, Ali to book the tickets.
a. promised b. thought of c. refused d. rejected
30. Woman: Can I try on this dress?
Assistant: Yes, madam. There is the changing-room.
Yes, madam. It looks wonderful.
Woman: O.K. I'll buy this.
In this dialogue, assistant her to buy the dress.
a. promised b. persuaded c. neglected d. intended
31. A: Don't forget to buy some coffee. You know, we have.....
coffee.
B: O.K. I'll write it down my shopping list.
a. stopped b. given up c. ended d. run out of
32. Customer: I paid 100.000TL for this radio but you didn't
give me a
Assistant: Sorry. I completely forgot, I was packing your radio.
a. receipt b. ticket c. deposit d. licence

APPENDIX B-1 (CONTINUED)

33. Receptionist : Good morning. Can I help you ?
 Patient : Yes, please. I'd like to for
 a medical examination.
 a. be an appointment | b. take an appointment
 c. make an appointment | d. do an appointment
34. A: Mete always looks happy, doesn't he?
 B: Yes. He is a person.
 a. mean b. cheerful c. clever d. sensible
35. A: Are you going to have a holiday in Marmaris?
 B: No. We a holiday in Marmaris.
 a. cancelled b. persuaded c. returned d. booked
36. Customer: The cassette recorder I bought last week won't
 work at all.
 Assistant : I wish to for that. Let me change it.
 a. enclose b. produce c. refuse d. apologise
37. Mother: Can you help me the washing-up?
 Daughter: O. K.
 a. do b. take c. make d. dish
38. Wife: You look sleepy and you're all the time. Why
 don't you wash your face?
 Husband: O.K. I'll do it.
 a. yawning b. feeling excited c. laughing d. feeling great
39. A: Did you see the doctor?
 B: Yes. He put me on a again.
 a. fat b. diet c. slim d. fit
40. A: For summer holiday, I would like to stay in a great big
 hotel with a disco, swimming pool and lots of people of my
 age.
 B: Do you mean somewhere ?
 A: Yes.
 a. quite b. out of date c. exciting d. peaceful

B

SINAV SÜRESİ 60 DAKİKADIR.

4 Yanlış cevap 1 doğru cevabı götürür.

-6-

ANADOLU UNIVERSİTESİ MUH.MİM.FAK.BAHAR DÖNEMİ
İNGİLİZCE I.ARA SINAVI:

Grub A

CHOOSE THE FOLLOWING CORRECT VOCABULARY FOR THE BLANKS:
(Boşluklar için en uygun kelimeyi seçiniz)

1. Receptionist: Good morning. Can I help you?
Patient: Yes, please. I'd like to for a medical examination.
a) be an appointment b) take an appointment
c) make an appointment d) do an appointment
2. A: Ramazan always looks happy, doesn't he?
B: Yes. He is a person.
a) mean b) cheerful c) clever d) sensible
3. A: Are you going to have a holiday in Marmaris?
B: No. He a holiday in Marmaris.
a) cancelled b) persuaded c) returned d) booked
4. Customer: The cassette recorder I bought last week won't work at all now.
Assistant: I wish to for that. Let me change it.
a) enclose b) produce c) refuse d) apologise
5. Mother: Can you help me the washing-up?
Daughter: O.K.
a) do b) take c) make d) dish
6. Wife: You look sleepy and you're all the time. Why don't you wash your face?
Husband: O.K. I'll do it.
a) yawning b) feeling excited
c) laughing d) feeling great
7. A: Did you see the doctor?
B: Yes. He put me on a again.
a) fat b) diet c) slim d) fit
8. A: For summer holiday, I would like to stay in a great big hotel with a disco, swimming pool and lots of people of my age.
B: Do you mean somewhere ?
A: yes.
a) peaceful b) quiet c) out of date d) exciting

SINAV SÜRESİ 60 DAKİKADIR.
4 yanlış cevap doğru cevabı götürür.

APPENDIX B-2 (CONTINUED)

-A-

9. Customer: I paid 100.000TL for this radio but you didn't give me a
- Assistant: Sorry. I completely forgot, I was packing your radio.
- a. receipt b. ticket c. deposit d. licence
10. A: Don't forget to buy some coffee. You know, we have coffee.
- B: O.K. I'll write it down my shopping list.
- a. stopped b. given up c. ended d. run out of
11. Woman: Can I try on this dress ?
- Assistant: Yes, madam. There is the changing-room.
Yes, madam. It looks wonderful.
- Woman: O.K. I'll buy this.
- In this dialogue, assistant her to buy the dress.
- a. promised b. persuaded c. neglected d. intended
12. Ali: Have you seen the movie at Ar1 Cinema ?
- Ahmet: What's on ?
- Ali: LAMBADA.
- Ahmet: O.K. Who is going to book the tickets ?
- Ali: Don't worry. I'll book the tickets.
- In this dialogue, Ali to book the tickets.
- a. promised b. thought of c. refused d. rejected
13. A: I have a pain in my arm.
- B: Why is that ?
- A: Because the nurse has just given me
- a. an injection b. a leaflet
c. some medicine d. some advice
14. Customer: How much does that radio cost ?
- Assistant: It is 250.000TL.
- Customer: It is too expensive for me. Would you please the price to 200.000TL.
- a. increase b. cut c. cheap d. borrow
15. Wife: Oh. We have lots of dirty clothes.
- Husband: We can have them washed by the washing machine.
- In this conversation, husband to use the machine.
- a. refused b. confused c. rejected d. offered

APPENDIX B-2 (CONTINUED)

PART: B-FIND THE GRAMMATICALLY CORRECT ANSWER:

(Gramer acısından dogru cevabı bulunuz)

- A -

16. Customs officer: Now, you should complete this form and write down
things: your name, age, address and occupation.
 Passenger: O.K. Sir. I'll do it.
 a) these b) those c) this d) that
17. A: You look ill today. What happened to you ?
 B: I'm not feeling well.
 A: You go out.
 a) had better not b) hadn't better not
 c) had better d) hadn't better
18. A: She doesn't eat fish.
 B:
 a) Adil does too b) So Adil does
 c) Nor does Adil d) So does Adil
19. We booked the holiday we paid a deposit of 100.000 TL.
 a) although b) because
 c) but d) and
20. A: Hello. Where are you going ?
 B: I'm going to the hairdresser to
 a) cut my hair b) cut me my hair
 c) have cut my hair d) have my hair cut
21. We went to a travel agency, we were planning a holiday.
 a) although b) because
 c) but d) and
22. A: Hello, Tom. Where are you going ?
 B: I'm going to apply for a new job. In fact,
 I for jobs for a month.
 a) am applying b) applied
 c) have been applying d) was applying
23. A: How did the exam go ?
 B: The questions were too difficult, so I answer
 some of them.
 a) could b) would
 c) should d) shouldn't

APPENDIX B-2 (CONTINUED)

-A-

24. Student: I went to bed late yesterday night and I
couldn't get up early, so I missed the exam.
I wish I
- a. didn't go to bed late b. hadn't got up early
c. had gone to bed early d. had got up late
25. Student: I didn't know what the vocabularies mean, so
I didn't answer the questions in the exam last
week.
Teacher: If you the vocabularies, you
answered the questions.
- a. had studied/would have b. studied/would have
c. had studied/would d. studied/would
26. Passenger: Excuse me sir. I've lost my passport, can I
get on the plane without passport?
Customs Officer: I'm sorry sir, you can't. You
forgot it.
- a. shouldn't b. should had
c. should have d. shouldn't have
27. A: We are going to Bursa. Are you coming with us?
B: I'm sorry, I'm not going with you because I have an
exam tomorrow. I wish I with you.
- a. had gone b. was going
c. am going d. can go
28. A: getting up early?
B: It makes me feel bored.
- a. What does it make b. How do you feel about
c. How does it make d. What makes you
29. A: Have you ever chopped onions?
B: Yes. Many times. But, it makes me want to
- a. laugh b. bored
c. cry d. amused
30. A: Fikret won the pools and he bought a big house in
the country.
B: I know that. But if I him, I buy
a small house near the sea.
- a. were/will b. was/would have
c. was/will d. were/would

APPENDIX B-2 (CONTINUED)

-A-

PART -C: READ THE PARAGRAPH and CHOOSE THE CORRECT ANSWER:

/Aşağıdaki soruları parçaya göre cevaplayınız/

Many people have begun to save energy. Paul Moore rides his bicycle to work to save gas. He began eight months ago. He carries his briefcase in the basket on the bicycle and heads for the office. He gets there in 30 minutes.

Paul has lost weight and become more healthy. He sleeps better at night. He enjoys the fresh air and sunshine. In bad weather he prefers bus and his car to his bicycle to go to work.

VOCABULARY:

ride: binmek

briefcase: çanta

head: ilerlemek, gitmek, yönelmek

basket: sepet

lose weight: kilo vermek

outdoors: açık hava

31. a. Paul leaves his briefcase home
b. He wants to save gas
c. He rides too fast
d. It takes an hour to get to the office.
32. He feels better now because -----
a. he drives too fast
b. he goes to the office on a bicycle
c. he carries the basket on the bicycle
d. he goes to the office on foot
33. In bad weather -----
a. he likes to go by bicycle
b. he prefers his bicycle
c. he goes by bus and his car
d. he enjoys riding his bicycle

Marcus and Beth Wilkins have finished their vacation.

They have spent two weeks in California. They have visited friends and relatives in many cities and towns. They have had a very good time. Today they have to go home. They have waited 40 minutes for a taxi. They have to get to the airport. Their plane leaves in 45 minutes.

VOCABULARY: vacation: tatil

relative: akraba

34. a. The Wilkins are going for a vacation
b. The Wilkins have started their vacation
c. The Wilkins have come to the end of their vacation
d. The Wilkins will start their vacation
35. a. They have to stay two weeks in California
b. They haven't enjoyed their vacation
c. They've waited many hours for a bus
d. They've visited many places and people
36. a. They're on their way home
b. They've to go to the bus station
c. Their train hasn't arrived yet
d. Their plane leaves in 2 hours

-A-

My name is June Taylor. My daughter Sara and I live in Fairbanks in Alaska. I'm a working mother. I work as a cameraman for a TV station in Fairbanks. My job is interesting, so I enjoy it. I didn't always work or plan to work, but I started to work ten years ago. This is my story. I married Charles Taylor in 1966 when we were very young. Charles was a salesman. He liked to move. That's why we moved to a new city every year. We lived in Chicago, Detroit and Miami. We lived far from our families. Charles wanted a large family and so did I. I didn't plan to get a job. I planned to stay home with children.

Sarah was born in Miami. Charles and I were very happy. I stayed home with Sara. We moved to Alaska the next year. I didn't have a job, but I had a hobby. I liked to take pictures of Sara and I learned to take very good pictures. I started to sell them, but it wasn't a real job. I always worked at home.

Charles suddenly became very ill in 1979. He had cancer and he died after six months. I had no money, no education and no friends. Then a TV station in Fairbanks saw my pictures. They asked me to work for them. I am very lucky to have the job, because I can earn enough money for Sara and me. We are independent and I don't worry about the future any more.

VOCABULARY:

salesman: pazarlamacı

cancer: kanser

move: taşınmak

die: ölmek

plan: planlamak

education: eğitim

take picture: fotoğraf çekmek

lucky: şanslı

suddenly: aniden

earn: kazanmak

become: olmak

independent: bağımsız

worry: üzülme

future: gelecek

37. Which sentence about June is false?

- a. She thinks she'll have a good future b. She moved to Fairbanks to find a job.
c. She enjoys her job because it is interesting
d. Her husband died of cancer about ten years ago.

38. June and Charles moved every year because -----.

- a. they wanted a large family. b. he wanted to move to different places.
c. he was a salesman and travelled d. they lived far from their families.

39. Before Charles became ill in 1979, June was -----.

- a. planning to work at the TV station b. working hard at the TV st.
c. staying home with the children d. taking some good pictures at home.

40. When June married Charles, she didn't plan to work. This means -----

- a. she didn't want to get a job b. she didn't try to get a job
c. she wanted to be a housewife d. she didn't need to get a job

APPENDIX B-3

The Results of the Midterm Test

SINAVA GIREN ÖĞRENCİ SAYISI : 76
 KAÇ YANLIŞ BİR DOĞRUYU GÖTÜRECEK : 4
 HER SCRUYA VERİLECEK PUAN : 2.500
 HERKESE EKLENECEK PUAN : 0.000
 ORTALAMA PUAN : 56.894
 EN DÜŞÜK PUAN : 8
 EN YÜKSEK PUAN : 88

P U A N D A Ğ I L I M I

8 İLE	24	ARASINDA	PUAN	ALANLARIN	SAYISI =	4
25 İLE	40	ARASINDA	PUAN	ALANLARIN	SAYISI =	8
41 İLE	56	ARASINDA	PUAN	ALANLARIN	SAYISI =	20
57 İLE	72	ARASINDA	PUAN	ALANLARIN	SAYISI =	27
73 İLE	88	ARASINDA	PUAN	ALANLARIN	SAYISI =	17

0 İLE	10	ARASINDA	PUAN	ALANLARIN	SAYISI =	2
10 İLE	20	ARASINDA	PUAN	ALANLARIN	SAYISI =	1
20 İLE	30	ARASINDA	PUAN	ALANLARIN	SAYISI =	6
30 İLE	40	ARASINDA	PUAN	ALANLARIN	SAYISI =	3
40 İLE	50	ARASINDA	PUAN	ALANLARIN	SAYISI =	14
50 İLE	60	ARASINDA	PUAN	ALANLARIN	SAYISI =	20
60 İLE	70	ARASINDA	PUAN	ALANLARIN	SAYISI =	13
70 İLE	80	ARASINDA	PUAN	ALANLARIN	SAYISI =	10
80 İLE	90	ARASINDA	PUAN	ALANLARIN	SAYISI =	7
90 İLE	100	ARASINDA	PUAN	ALANLARIN	SAYISI =	0

APPENDIX B-3 (CONTINUED)

L.NCI SORU	T E S T		G R U B U		IÇIN E	TOPLAMLAR:	
	A	B	C	D		BOS	DCEVAP
1	1	9	25	2	0	0	C
2	0	36	1	0	0	0	B
3	28	0	8	1	0	0	A
4	7	7	7	14	0	2	D
5	24	1	6	6	0	0	A
6	29	2	2	3	0	1	A
7	1	36	0	0	0	0	B
8	3	2	1	29	0	2	D
9	23	0	9	4	0	1	A
10	3	14	12	7	0	1	D
11	10	23	1	2	0	1	B
12	19	3	7	1	0	7	A
13	20	3	6	4	0	4	A
14	12	18	5	1	0	1	B
15	4	3	5	18	0	7	D
16	19	3	8	6	0	1	A
17	13	1	7	16	0	0	A
18	4	2	20	10	0	1	C
19	10	3	0	24	0	0	D
20	21	0	13	2	0	1	D
21	3	32	0	2	0	0	B
22	1	2	30	2	0	2	C
23	15	3	1	16	0	2	A
24	16	0	21	0	0	0	C
25	23	8	3	2	0	1	A
26	7	2	4	23	0	1	D
27	6	15	4	11	0	1	B
28	1	28	6	2	0	0	B
29	4	2	25	4	0	2	C
30	6	5	1	25	0	0	D
31	9	19	0	6	0	3	B
32	1	29	4	1	0	2	B
33	0	7	26	3	0	1	C
34	0	0	35	0	0	2	C
35	21	0	0	15	0	1	D
36	29	1	1	2	0	4	A
37	7	16	5	7	0	2	B
38	4	10	16	6	0	1	C
39	4	7	10	12	0	4	D
40	3	4	15	10	0	5	C

APPENDIX B-3 (CONTINUED)

2. NCI SORU	T E S T		G R U B U		IÇİN E	TOPLAMLAR:	
	A	B	C	D		BOS	DCEVAP
1	10	0	28	0	0	1	C
2	30	5	1	2	0	1	A
3	5	1	1	32	0	0	D
4	5	19	4	10	0	1	B
5	2	32	3	2	0	0	B
6	4	1	30	3	0	1	C
7	4	4	1	30	0	0	D
8	22	2	3	12	0	0	A
9	1	1	37	0	0	0	C
10	2	34	1	1	0	1	B
11	21	0	11	6	0	1	D
12	4	4	0	29	0	2	D
13	2	4	26	5	0	2	C
14	12	4	6	16	0	1	A
15	26	8	0	4	0	1	A
16	27	3	6	0	0	3	A
17	33	1	1	2	0	2	A
18	3	0	0	35	0	1	D
19	0	39	0	0	0	0	B
20	1	0	26	11	0	1	C
21	35	0	1	1	0	2	A
22	16	7	5	10	0	1	A
23	13	3	17	3	0	3	A
24	6	2	18	10	0	3	C
25	4	10	3	17	0	5	D
26	1	4	1	30	0	3	D
27	6	24	4	4	0	1	B
28	32	1	2	0	0	4	A
29	20	4	6	6	0	3	A
30	1	32	2	0	0	4	B
31	3	10	16	10	0	0	D
32	28	2	4	2	0	3	A
33	2	8	29	0	0	0	C
34	0	36	2	1	0	0	B
35	28	2	5	3	0	1	A
36	7	3	13	15	0	1	D
37	23	3	9	2	0	2	A
38	34	0	1	3	0	1	A
39	1	36	0	2	0	0	B
40	0	0	34	2	0	3	C