Durham University

Durham E-Theses

Testing English as a foreign language: a case study of classroom tests in Qatar

Tariq Mustafa Abbara

How to cite:

Abbara, Tariq Mustafa (1991) Testing English as a foreign language: a case study of classroom tests in Qatar. Masters thesis, Durham University.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a https://etheses.durham.ac.uk/id/eprint/5989/ is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
https://etheses.durham.ac.uk

Testing English as a Foreign Language: a Case Study of Classroom Tests in Qatar

by

Tariq Mustafa Abbara

A thesis submitted for the degree of Master of Arts

School of Education
University of Durham
England

The copyright of this thesis rests with the author. No quotation from it should be published without his prior written consent and information derived from it should be acknowledged.

1991

- 8 SEP 1992

"One assumption is that change does not occur unless the particulars of a school and its context are taken into account. A second is that a school staff will not be committed to a change effort unless they have had the opportunity to be involved in decisions concerning the scope of the project.

(David Hopkins, 1986, P. 7)

In the name of Allah, The Beneficient,

the Merciful.

Dedication

I humbly dedicate this work to my father, Mustafa, my wife, Enayah and my children: Linda, Danyah, Mohammad, Ahmad, Mahmmoud, Hamza and Hussam with love.

Abstract

The thesis is concerned with testing English as a foreign language in general and concentrates on testing in Qatar in particular.

Chapter I provides a brief overview of education in Qatar to form a solid basis for the study.

Chapter II presents an overview of the historical stages of development of testing and relates the Qatari situation to that framework.

Chapter III is devoted to the different kinds of test and their advantages and disadvantages and relates this framework to the Qatari situation by describing the kinds of test currently in use in Qatar.

Chapter IV focuses on the issues of reliability and validity. These two qualities are dealt with as a basis for judging and improving achievement tests.

Chapter V will focus even more closely on the kinds of tests used in Qatari schools. These tests are written either by the Inspectorate or by classroom teachers to reflect the nature of the course. This chapter describes a case study which is based on "the Crescent English Course". It also provides a full description of the materials being used as well as the examination frequency and the allocated marks.

Chapter VI is a summary chapter which deals with recommendations to improve the current status of classroom tests.

Acknowledgement

First of all, I must express my sincere and thankfulness to Allah, without whose help this work would never have been possible.

I also would like to express my deepest appreciation and sincere gratitude to my supervisor, Dr. Michael Byram, Director of Higher Degrees at the University of Durham for his help and wise counsel. I am deeply indebted to him for patience and invaluable advice throughout all the stages of this work.

Thanks must also go to Mr. R. F. Goodings for his encouragement and the library staff at Durham University, Aston University in Birmingham, and Qatar University for their help.

I am also wholeheartedly grateful and obliged to Mr. Hussien Mahmoud Abbara, Educational Officer, in the office of the cultural attache (Embassy of Qatar) in London for his inclinable support and encouragement.

Last, but not least, I would like to express my sincere gratitude to my wife without whom I would never have been able to continue and complete this work.

C O N T E N T S

Page

DEDICATION.....i

ABSTRACT.....ii

ACKNOWLEDGMENT.....iii

CONTENTS.....iv

CHAPTER 1. BRIEF OVERVIEW OF EDUCATION IN QATAR.....

Topical Outline.....1

1.0 Introduction.....3

1.1. Brief Overview of Education in Qatar.....4

1.2. History of English Language Teaching in Qatar..9

1.2.1. Research Stage.....11

1.2.2. The Stage of Writing the Materials.....12

1.2.3. The Stage of Training the Teachers.....12

1.2.4. The Objectives of the Crescent Course.....17

1.2.4.1. Listening.....18

1.2.4.2. Speaking.....19

1.2.4.3. Reading.....20

1.2.4.4. Writing.....20

1.3. Factors Affecting Standards of Achievement.....22

1.3.1. Teacher Qualifications.....22

1.3.2. Learner Motivation.....28

1.3.3. Textbooks.....30
1.3.4. Tests.....32
1.4. The System of Testing in Qatar.....34
1.4.1. The Testing Scheme.....36
1.4.2. Examination Times.....37
1.5. Problems of Testing.....44
1.6. The Need and Origin of the Study.....50

CHAPTER 2. OVERVIEW OF HISTORY AND DEVELOPMENT OF LANGUAGE TESTING.....

Topical Outline.....53
2.1. Introduction.....54
2.2. History of Language Testing.....54
2.2.1. The Early Intuitive Stage (before 1850).....55
2.2.2. The Pre-Scientific Stage (before 1850-1950/60)..56
2.2.3. From 1900 to 1915.....60
2.2.4. from 1915 to 1930.....61
2.2.5. From 1930 to 1945.....62
2.2.6. From 1945 to 1960.....62
2.2.7. Modern Testing.....63
2.2.8. Testing in the 1980s.....65
2.2.8.1. The First Characteristic of Current Testing....66
2.2.8.2. The Second Characteristic.....67
2.2.8.3. The Third Characteristic.....68
2.2.8.4. The Fourth Characteristic.....69
2.3.1. The Historical Development of Testing in Qatar.70

CHAPTER 3. PURPOSES AND KINDS OF TESTS.

Topical Outline.....76
3.0. Introduction.....79
3.1. The Significance of Testing in the Classroom.....80
3.2. The Purposes of Testing in the Classroom.....83
3.2.1. Measuring Outcomes of Instruction.....84
3.2.2. Improving the Curriculum.....85
3.2.3. Placement and Grouping.....86
3.2.4. Predicting Future Performance.....88
3.2.5. Describing Realistic Standards of Achievement....89
3.2.6. As a Diagnostic Procedure.....89
3.2.7. To Grade Students for Promotion.....90
3.3.1. Different Types of Test.....92
3.3.2. Achievement Testing.....94
3.3.3. Standardized Tests.....100
3.3.4. Teacher-Made Tests.....101
3.3.4.1. Advantages of Teacher-Made Tests.....103
3.3.4.2. Disadvantages of Teacher-Made Tests.....104
3.3.5. Norm-Referenced Testing.....106
3.3.5.1. Advantages of Norm-Referenced Tests.....109
3.3.5.2. Disadvantages of Norm-Referenced Tests.....110
3.3.6. Criterion-Referenced Testing.....110
3.3.6.1. When to Use CR Tests.....114

3.3.6.2. Criterion-Referenced Test Construction.....118
3.3.6.3. Tasks for Level 1.....119
3.3.6.3.1. Listening.....119
3.3.6.3.2. Speaking.....120
3.3.6.3.3. Reading.....121
3.3.6.3.4. Writing.....122
3.3.6.4. Characteristics of CR Testing.....126
3.3.6.4.1. Advantages of CR Tests.....130
3.3.6.4.2. Disadvantages of CR Tests.....132
3.3.7.1. Differences Between CR and NR Tests.....133
3.3.8.1. Objective Testing.....136
3.3.8.2. Advantages of Objective Testing.....137
3.3.8.3. Disadvantages of Objective Testing.....138
3.3.8.4. Types of Objective Testing.....138
3.3.8.4.1. True-False Items.....139
3.3.8.4.2. Advantages of True-False Items.....139
3.3.8.4.3. Disadvantages of True-False Items.....140
3.3.9.1. Matching Items.....141
3.3.9.2. Advantages of Matching Items.....142
3.3.9.3. Disadvantages of Matching Items.....143
3.3.10.1. Multiple-Choice Items.....143
3.3.10.2. Advantages of Multiple-Choice Items.....144
3.3.10.3. Disadvantages of Multiple-Choice Items.....145
3.3.11.1. Subjective Testing.....145
3.3.11.2. The Essay Test.....146

3.3.11.3. Scoring the Essay Test.....148
3.3.11.4. Advantages of the Essay Test.....148
3.3.11.5. Disadvantages of the Essay Test.....149
3.3.12.1. Test Types in Qatar.....150

CHAPTER 4. VALIDITY AND RELIABILITY-FUNDAMENTAL QUALITIES OF GOOD TESTING.....

Topical Outline.....153
4.1. Introduction.....154
4.2. Validity.....155
4.2.1. Literature Survey.....155
4.2.2. Types of Validity.....158
4.2.2.1. Face Validity.....158
4.2.2.2. Content Validity.....160
4.2.2.3. Construct Validity.....164
4.2.2.4. Predictive Validity.....167
4.3. Reliability.....171
4.3.1. Introduction.....171
4.3.2. Definition of Reliability.....171
4.3.3. Ways of Estimating Reliability.....174
4.3.3.1. Test-retest Technique.....174
4.3.3.2. Alternative Forms of the Original Test.....176
4.3.3.3. Split-half Technique.....178
4.4. Relationship Between Reliability and Validity179

CHAPTER 5. CASE STUDY BASED ON "THE CRESCENT ENGLISH COURSE"..

Topical Outline.....182
5.0. Intoduction (course background).....183
5.1. Course Objectives.....185
5.2. Analysis of Three Teacher-Made Tests.....187
5.3. Test One.....188
5.3.1 Testees' Background.....188
5.3.2 Test Specification.....188
5.4 Test Two.....194
5.4.1 Testees' Background.....194
5.4.2 Test Specification.....194
5.5 Test Three.....196
5.5.1 Testees' Background.....196
5.5.2 Test Specification.....196
5.6. Implementation for Future Practice.....201

CHAPTER 6 CONCLUSION AND RECOMMENDATION.....

6.1 Introduction.....206
6.2 Conclusions.....209
6.3 Recommendations.....211

REFERENCES.....222
BIBLIOGRAPHY.....229

APPENDICES.....245

Appendix A 246

Primary Level Objectives.....246

Appendix B 249

Preparatory Level Objectives.....249

Appendix C 252

Secondary Level Objectives.....252

Appendix 1A 255

Description of The Course Materials.....255

Appendix 1B 256

Primary Six 256

Appendix 2 257

Preparatory One 257

Appendix 3 258

Preparatory Two 258

Appendix 4 259

Preparatory Three 259

Appendix 5 260

Secondary One 260

Appendix 6 261

Secondary Two 261

Appendix 7 262

Secondary Three 262

Appendix 8 263

Description of Course Contents.....263

Appendix 9 264

Marks Allocation-Monthly Exams (oral).....264

Appendix 10 265

Marks Allocation-Monthly Exams (written).....265

Appendix 11 266

Marks Distribution on Written Exams (all levels)..266

Appendix 12 267

Model Examination (primary level).....267

Appendix 13 273

Model Examination (preparatory level).....273

Appendix 14 280

Model Examination (secondary level).....280

Chapter One

BRIEF OVERVIEW OF EDUCATION IN QATAR:

Topical Outline

1.0 Introduction.

1.1 Brief overview of education in Qatar especially EFL.

1.2 The history of ELT in Qatar.

1.2.1. Research stage.

1.2.2. The stage of writing the materials.

1.2.3. The stage of training the teachers

1.2.4. The objectives of the Components of the English Syllabus Crescent English Course (CEC).

1.2.4.1. Listening.

1.2.4.2. Speaking.

1.2.4.3. Reading.

1.2.4.4. Writing.

1.3 Factors affecting standards of achievement in EFL.

1.3.1. Teacher qualification.

1.3.2. Learner motivation.

1.3.3. Textbook.

1.3.4. Tests.

1.4. The system of testing.

1.4.1. Scheme of the testing system.

1.4.2. Examination times.

1.5 Problems of testing.

1.6 The need and origin of the thesis.

Chapter One:

1.0 Introduction:

Complaints have been repeatedly expressed by both the Ministry of Education and teachers of EFL in QATAR about the low standards of students' attainment in English although time, money, and energy have been spent in the form of a new approach of teaching and new materials that take the students' environment into account, but unfortunately the failure continues to exist. Dr. Neil Bratton, who used to be the English language consultant in Qatar, wrote a report in 1983 in which he says:

> Why do we continue to fail even after making the greatest efforts to improve the language component which includes approaches, methods, techniques, syllabuses, texts, aids and tests ? Is it the fault of the curriculum developers who are responsible for the above improvement? Is it the fault of the learners who are lazy and unmitivated? Is it the fault of the teachers who are weak in English and inadequately trained? or does the fault lie in the system within which these three components operate. (P. 1)

In this chapter, I shall concentrate on the Qatari situation to provide a clear idea about the coming scene. I shall introduce a brief history of TEFL in Qatar. The problems of standards of achievement in ELT will be considered to

diagnose the difficulties in the present situation. The system of testing will be looked at in order to explain the problems of the current procedures of testing. Finally, the need for the study will be explained to reinforce the significance and scope of the thesis.

1.1 Brief Overview of Education in Qatar Especially EFL:

The state of Qatar is situated halfway along the western coast of the Arabian Gulf. It consists of a peninsula projecting northwards about 150 Km. into the Gulf and covers an area of 11.437 square Km. The state of Qatar has a population of 235.000, 80% of whom live in Doha, the capital city. Arabic is the official language, although English is widely spoken.

The ruling family is that of the Al-Thani, which settled in Qatar around the middle of the nineteenth century. Qatar's ruler is HH Sheikh Khalifa Bin Hamad Al-Thani. The state of Qatar became independent on September 3, 1971. Qatar is an active member of the Gulf Cooperation Council (GCC), The Arab league, the non-aligned movement and the United Nations.

Oil, both onshore and offshore, is the backbone of Qatar's economy. The first oil production in commercial quantities was in 1949. Full ownership of the oil was achieved in 1977. Qatar joined the Organization of Oil-Exporting Countries (OPEC) in 1961 and participated as an active state in the foundation of

the Organization of Arab Oil-Exporting Countries (OAPEC) in 1970. The discovery of a huge gas field, in the north of the country, will guarantee the welfare of the Qatari people in the future. This gas field is one of the largest gas fields in the world.

Education in Qatar is free of charge at all levels, including those of the university and post graduate study. There are 95 schools for boys and 96 for girls at Primary, Preparatory, Secondary and Specialized fields of study such as Industrial and Commercial schools. The grand total of enrollment is 61914 students at all levels, (19274 boys, 30614 girls) 54 % of them are Qatari students. (see table 1).

Table 1

Number of schools, classes and students in Qatar. (source, Ministry of Education= Annual Report, 1990)

		Primary			Preparatory			Secondary			Specialized				Grand Total		
		Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total	Reli.	Com.	Indu.	Total	Boys	Girls	Total
Number of Schools		55	49	104	23	27	50	14	20	34	2	1	1	4	95	96	192
Number of Classes		690	657	1347	250	256	506	151	206	357	16	5	13	34	1125	1119	2244
Students' Nationality	Qatar	10035	9391	19426	4126	4344	8470	1921	3017	4938	121	91	396	608	16090	16752	33442
	Non Qatar	9239	8780	18019	3173	3117	6290	1879	1965	3844	297	17	5	319	14610	13862	28472
	Total	19274	18171	37445	7299	7461	14760	3800	4982	8782	418	108	401	927	19274	30614	61914
Qatari Ratio		52%	52%	52%	57%	58%	57%	51%	61%	56%	29%	84%	99%	66%	53%	54%	54%

In 1973, Qatar University was opened with two teacher training colleges. In 1989/1990, 659 students have graduated from Qatar University, 535 of them were girls. Qatari post-graduates of both sexes pursuing their studies in Gulf States, Arab States, U.S.A. and other countries are 918 students. [Table 2] summarizes, according to the information available from the Ministry of Education, the number of students who are pursuing their studies in different countries.

Table 2

Post Graduate Studies + Boys And Girls (Only Qatari Students)

Country	Gulf States		Arab States		U.SA. @ Austra		Europe		Total		Grand Total
Field of Study	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	
Post Graduate	11	17	49	97	73	29	44	43	177	186	363
Medicine	29	107	10	9	-	-	50	-	89	116	205
Medical Science	-	14	-	3	4	2	-	-	4	19	23
Computer @ Maths	-	1	2	-	8	5	1	-	11	6	17
Engineering	18	-	5	-	100	-	2	-	125	-	125
Religion/Law	4	-	32	2	1	-	-	2	37	4	41
Social Studies	-	1	5	-	-	3	-	-	5	4	9
Languages	-	1	1	-	-	4	5	1	6	6	12
Political Science	-	2	17	1	47	3	-	-	64	6	70
Education/Psycho.	-	1	-	1	-	1	-	2	-	5	5
Art Education	2	1	8	-	-	-	-	1	12	2	14
Others	1	-	8	-	5	-	-	-	14	-	14
Special Cases	17	3	-	-	-	-	-	-	17	3	20
Grand Total	82	148	137	113	239	47	103	49	561	357	918

The educational system is divided into three levels: Primary (6 years), Preparatory (3 years) and Secondary (3 years). The English language is taught from primary 5. The school year runs from September 15th to June 15th, divided into two semesters by a two-week break which usually occurs at the end of January / beginning of February.

1.2 History of English Language Teaching in Qatar:

1.2.1 Teaching English in the Qatari schools began as in any other Arab country, particularly in the Gulf States, by adopting the Structural Approach which first began in Europe and then spread to other parts of the world as a reaction against the Grammar-Translation Approach as a result of the expansion in foreign language teaching and the increasing contact between different communities. The Structural Approach was represented in a series of books, which was dominant in the late 1960s and the first half of the 1970s. The books that were in use in Qatar at that time were called: Living English For The Arab World. by W. S. Allen and Ralph Cook. This series was used in Qatar in 1965 as a result of its success in Kuwaiti schools which began to adopt the same series in 1962-3.

When Qatar began to send students to pursue their higher studies in many European and American Universities, the students were in need of English courses to raise their

standards to cope with the new world of study at foreign Universities on the one hand, and the dissatisfaction which was expressed by both teachers and students against the Structural Approach lead to a change from the Structural to the Communicative Approach. Qotbah (1990) lists four facts that have lead to the abandonment of the Structural approach.

1. The syllabus neglected to a great extent listening and speaking skills. Dialogues for practising speaking were artificial, very long and hard to remember or practice.
2. Textbooks were full of repetitive, boring drills in each unit and students were always asked to memorize lists of words in order to pass the exam.
3. Subjects were often outmoded and uninteresting for the students; for example there were some long stories full of difficult words. In addition, textbooks did not have many pictures to interest the students but looked more like novels.
4. The textbooks contained some very long words which were difficult for the students to memorize or even to pronounce. In short a new syllabus was needed to cope with the development of language learning theories in an effective way. (P. 24)

In 1975, the Ministry of Education made contact with the American University in Beirut. In April 1975, a team was appointed to produce a comprehensive study of the existing textbooks. As a result of this study, the team diagnosed the following reasons for the students' poor standards.

1-Poor understanding of the importance of objectives
2-Teachers are not well qualified in teaching English.
3-The teaching methods are inadequate.
(Arabic report about TEFL from 1975 to 1988, Ministry of Education)

The same study suggested a way for modification which had three stages:

1.2.1 Research stage:

At this threshold level, the learners' needs for the language in the future were identified. To achieve these needs, the Ministry did the following:

a) An English Proficiency Test was imposed on all teachers of English to decide their ability to use English in classrooms.

b) A sample group of students, including the students of the Faculty of Education in Qatar University, took the same test.

c) A questionnaire was given to all those who took the Proficiency Test.

d) Uses of English in real Qatari life were investigated through another questionnaire which was given to young employees to define how and when they use English in real life situations.

1.2.2 The Stage of Writing the Materials:

At this stage the kind of materials was identified due to the results of the two previous questionnaires. At the end of 1975, an agreement was signed between the Ministry of Education and Oxford University Press to write and publish the textbooks. In 1976, the first Crescent materials were used on an experimental basis in Qatar. The materials expose the learners through the written work and the the recorded materials, to a wide variety of authentic English.

1.2.3 The Stage of Training the Teachers

At this stage, both prospective and experienced teachers were trained in how to use the newly introduced

materials. The Crescent English Course came into being at this stage. The basic approach to curriculum reform in ELT for Qatar was developed in a series of workshops and seminars organised by the Oxford University Press in Qatar (June 1976), London (1976), Abu Dhabi (1977), Cairo (January 1977). These workshops and seminars were prepared to train members of the English Language Development Centre (see below). Those who were trained by Oxford University Press (OUP) were appointed to train prospective teachers every new school-year.

The aim of the Crescent Course is to develop in pupils the communicative competence necessary to use English in real-life situations. Then, in 1975, the English Language Development Centre was founded to supervise the project. The following people were appointed to direct the English Language Development Centre:

Dr. Neil Bratton	: English Language Consultant
David Kirwan	: Teacher Trainer
Helen O'Neill	: = =
Alan Swales	: Assistant Teacher Trainer
A. Abu Jalala	: = = =
Margaret Aitken	: = = =
S. Shafi	: = = =
David Aitken	: Testing Co-ordinator

In 1978, the phased introduction of the materials into all schools began and the Ministry appointed two in-service teacher-trainers. In the meantime, further seminars were held in Abu Dhabi and Cairo. The Crescent English Course is now used at all levels (Primary, Preparatory, and Secondary) in all schools in Qatar, the United Arab Emirates, Kuwait, and on an experimental basis in Bahrain.

The English Language Development Centre was originally founded for two reasons. The first was to supervise the newly adopted approach, and the second was to train teachers in how to deal with the new materials. In-service training is usually combined with a teacher's school schedule. In Qatar, some teachers had specially arranged schedules so that they could attend in-service training. In fact, there were many reasons for in-service training. Some of these reasons are the following:

1. To familiarize teachers with the newly adopted procedures of teaching.
2. Orientation to newly used materials.
3. Improving teachers' proficiency in English.

The use of the Crescent English Course in Qatar, required widespread efforts to orient teachers, both new and experienced, to the philosophy and methods of dealing with the

new materials. Such training courses were of great importance as a logical result of the change from Structural to Communicative approach. Teachers were in need of these courses to understand how to use the " functional " English methodology in the new world of teaching. Since teaching is a highly creative profession, it demands the introduction of new ideas and thoughts. Teachers cannot invent new techniques by themselves to the new field of teaching. Instead, they are in need of orientation and refresher courses from time to time to develop new professional skills in the target domain. The training programmes that were used in Qatar did not achieve the goal because there was not enough time to concentrate on the two important variables in these programmes. These two variables are the teacher who is the key person in the teaching-learning process and the situation of teaching new materials which the teacher is being prepared for. When we want to train teachers on a new area, we should consider the difficulties that we may face. In the Qatari situation, the difficulties lie in the sharp turning from the traditionally adopted approach (Structural) to newly adopted one (Communicative). The programmes should aim at changing the teacher activities from teaching in a teacher-centred classroom to a student-centred one.

In a word, the training courses that were designed in Qatar failed to achieve the intended purposes through a fault

which was not related to the Ministry of Education, but it was a matter of adequate time.

The Crescent Course adopts a communicative approach which has three main implications.

1-We must teach the language not the rules.
2-The learning should take place in as natural situations as possible
3-Enjoyment, as an essential motivating factor, becomes central

The Crescent English Course seeks to provide students with opportunities for a variety of activities for communicating with one another and with their teacher in a meaningful and natural way. This is the first important aspect of the communicative approach. The second important aspect is that learning a language is an individual process. Almost all the language a child hears and speaks is of importance to him/her as an individual in a social context. The course has been designed to provide students with social situations through the adopted technique of group work, which enables students to learn indirectly by co-operating with one another through language. Lessons are developed, as far as possible, around activities which are motivating in themselves and which reflect the interest of students. Examples of this are the use of

songs, games and comic strip stories.

The native language of the learners is sometimes used at the early stages of learning when comprehension is at a complete standstill and the teacher has honestly made every possible effort to explain whatever needs explaining in the target language. There is nothing as frustrating to students as feeling completely left out of an entire lesson because they did not understand something at the beginning of it.

1.2.4 The objectives of the Crescent Course were not set in advance so that teachers can make every possible effort to achieve them effectively. The Crescent Course materials were written without a clear idea about its objectives. This was a major criticism of the course. In a later time, in 1985, the course objectives were produced on a formal basis by the Ministry of Education. These objectives include listening, speaking, reading and writing for each of the three levels. (Primary, Preparatory and Secondary). Let us quote what has been issued by the English Inspectorate in 1985 to give an example of these objectives. The objectives I quote here are meant for the Preparatory level. They are:

1.2.4.1 Listening:

By the end of the preparatory level, the students should be able to do the following:

1- Listen to orders and instructions and carry out a talk (playing a game, filling in a form, translating directions etc.)

2-Listen to authentic English transmitted through the media or airport announcements or a telephone conversation, and be able to extract specific information.

3-Listen and understand short extracts from the media like the news, commentaries, and plays.

4-Listen to and identify three basic themes of a short conversation between two speakers, as well as their opinions and kind of relationship they have (friendly or otherwise).

5-Understand the main idea and important points of a story or a dialogue.

6-To identify the situation from the context of a conversation between two people (a visit to the doctor, buying something from a store etc.).

1.2.4.2. Speaking:

By the end of the Preparatory level, the pupils should be able to express themselves in a clear and acceptable way in different situations such as:

1-Taking part in a short social conversation concerning home, family, school, travel, media and local environment.

2-Coping with English language contact situations such as helping a foreigner in the market or the street; giving him simple directions; giving simple explanations of situations related to the local environment.

3-Handling language situations in a foreign country where English is the means of communication.

4-Taking part in a telephone conversation to give an invitation, make an apology or question a certain personal or practical situation.

5-Telling a story or describing an accident or relating an incident.

1.2.4.3 Reading:

By the end of the Preparatory level, the pupils should be able to read and understand the following:

1-A short story or expository passage in simple English with only a few unknown words, making an intelligent guess at the meaning of unknown words from their context.

2-Extract information from a letter or an article and act accordingly.

3-Interpret and follow written instructions and act accordingly.

4-Understand the news, reports and advertisements.

5-Extract information from an alphabetical list (telephone directory) and interpret maps, diagrams, tables and graphs.

6- Use simple dictionaries.

1.2.4.4 Writing:

1-Pupils should be able to write about themselves, their families, friends and school.

2-They should be able to summarise a spoken story and write descriptions of incidents they have witnessed.

3-They should be able to write letters (e.g. to a pen-friend), asking for information or advice.

4-They should be able to fill in forms such as for passports or at the airport.

5-They should be able to write down a simple summary of information from a map or a diagram.

Stating the instructional objectives in advance by the Ministry of Education in Qatar helps teachers to describe in a relatively specific manner what a student should be able to do or produce, or they describe the characteristics that a student should possess at the end of a course of study. Teachers are invited to pursue practical strategies in testing the instructional objectives they deal with during the academic year. These strategies can be formulated through cooperation between colleagues to ease the work-load associated with the task. What happens at the Qatari schools is exactly the opposite. The objectives are written and distributed to the teachers at the beginning of the school year. Then teachers read these objectives and put them aside. No careful thought seems to be given to the ways of testing these objectives to know how many of them have been achieved. Another point which is still dominant among teachers in Qatari schools is that the instructional objectives they receive from the Ministry of Education are unquestionable. On the contrary, there is no need

for teachers to limit themselves to the stated objectives. Although the objectives are always set in advance, teachers should study them carefully to decide what strategies should be used to test these objectives. At this stage cooperation between colleagues plays an important role if teachers wish to act seriously.

1.3 Factors Affecting Standards of Achievement:

Most educators in Qatar are not satisfied with the present state of students' achievement in English as a foreign language. They always blame teachers and consider them responsible for the poor standards in English. It must be always kept in mind that teachers are not the only people to be blamed, rather there are so many factors affecting standards of achievement in classrooms. Among them are the following:

1.3.1 Teacher qualifications:

It is vitally important that teachers of English as a foreign language should be adequately prepared for their future task. In Qatar, some teachers are well-trained in teaching the language, but not in testing it. The majority are only introduced to a course which usually lasts only two weeks at the beginning of each academic year. This course of orientation

is always intended for inexperienced and prospective teachers before they face their first teaching experience. This time is not enough to familiarize teachers with the textbooks they will teach. The result is that many teachers have been poorly and inadequately prepared, often through no fault of their own. It is clear that teaching by persons who cannot meet the minimal standard of qualification (the knowledge of effective methods and techniques of language teaching) will not make a distinctive contribution to language learning and testing. Preparation of teachers for their inescapable responsibility of testing their students should begin at the pre-service level. A specialist in test construction could be of considerable assistance to such a group of teachers.

Broadly speaking, teachers of English in primary and preparatory schools can be divided into four categories on the basis of their academic and professional qualifications: (tables 1, 2). Table (1) shows us that teachers who teach English in Primary schools are divided into two categories according to their qualifications:

(A). Teachers who hold a university degree or higher than university degree are only 45 teachers out of 136 in Primary level.

(B). Teachers who hold lower than university degree are either Diplomas holders (two years after the General

Secondary Certificate) or only General Secondary Certificate .

Table (2) indicates that teachers who teach English in Preparatory schools are divided into two categories as far as qualifications are concerned:

(A). Teachers who hold a degree lower than B.A. are 9 out of 176 which is the grand total. They hold diplomas, two years study in teacher-training institutes after the General Secondary Certificate.

(B). Teachers who hold B.A. are also divided into two sub categories:

1. B.A. (non-educational) in the English language and its literature.
2. B.A.(educational) in teacher training and education.

Table 1

Number of men and women teachers in Primary schools in Qatar and their qualifications

Teachers	B.A. and higher (A) Educational	B.A. and higher (B) Non-Educational	Lower than B.A. (A) Educational	Lower than B.A. (B) Non-Educational	Grand Total	NO.of Qatari
Islamic Religion	116	122	118	47	403	273
Arabic Language	111	094	080	19	304	139
English Language	045	026	061	04	136	027
Mathematics	091	049	114	15	269	027
Science	059	096	068	11	234	136
Social Studies	065	053	021	06	145	100
Art Education	031	030	044	13	118	049
Physical Edu.	081	022	020	08	131	028

Data Keys

B.A. & Higher for English Language.
Lower than B.A.
(A)-Teachers with Diplomas from teacher training institutes.
(B)-Teachers with General Secondary Certificate.

Source : Ministry of Education, Qatar, Annual Report, [1989-1990]

Table 2

Number of men and women teachers in Preparatory schools and their qualifications in Qatar.

Teachers	B.A. and higher (A)	(B)	Lower than B.A. (A)	(B)	Grand Total	NO. of Qatari
	Educational	Non-Educational	Educational	Non-Educational		
Islamic Religion	062	087	03	02	154	74
Arabic Language	104	106	—	01	211	69
English Language	124	043	09	—	176	60
Mathematics	135	023	04	—	162	20
Science	074	048	02	—	124	56
Social Studies	111	043	01	—	155	86
Art Education	036	032	03	—	072	20
Physical Education	054	—	04	—	058	05

Data Keys

B.A. & Higher for English Language.
Lower than B.A.
(A)-Teachers with Diplomas from teacher training institutes.
(B)-Teachers with General Secondary Certificate.

Source : Ministry of Education, Qatar, Annual Report, [1989-1990]

Table 3

Number of men and women teachers in Secondary schools in Qatar and their qualifications

Teachers	B.A. and higher (A)	(B)	Lower than B.A. (A)	(B)	Grand Total	NO. of Qatari
	Educational	Non-Educational	Educational	Non-Educational		
Islamic Religion	036	071	—	—	107	33
Arabic Language	048	105	—	—	155	08
English Language	086	053	—	—	139	28
French Language	003	011	—	—	014	—
Mathematics	076	038	—	—	114	—
Biology	029	020	—	—	049	21
Geology	029	021	—	—	049	10
Chemistry	031	021	—	—	052	15
History	027	023	—	—	050	17
Philosophy	015	037	—	—	050	23
Geography	039	013	—	—	052	24
Art Education	016	012	01	01	030	01
Physical Education	021	001	—	—	027	01

Data Keys: as in table 1 & 2.

While in Secondary schools, table 3 gives the impression that teachers at this stage are in a better situation. All teachers have university degrees, either in education or in the English language and its literature, those teachers who studied the English language at the university level have not undergone the necessary training at the university. Special attention should be given to them when they intend to teach and test English. We run the risk of turning a promising teacher away from his chosen profession if his early teaching efforts are met with poor results and consequent frustration. In reality, there are very few M. Ed. or Ph. D. holders teaching in the public schools. The essence of the matter is that most English teachers have B. A.'s in areas other than English Language Teaching from Arab Universities, where there are very few courses in TEFL methodology.

1.3.2 Learner motivation:

Students in Qatar lack the adequate motivation to learn English. Dr. Qotbah, 1990 quotes Widdowson 1983 to point out the result of students' lack of motivation saying:

> A lack of motivation on the part of the students may arise either from a rejection of the aims presupposed by the objectives, or from a

rejection of the objectives as a valid mediation towards aims that they do not accept. (P. 31)

I think the classroom teacher can play an important role in raising the students' motivation in the classroom. This can be done on the basis of convincing them that the language they learn will be of great importance to them when they decide to pursue higher studies, conduct a business, or take a trip to a foreign country. On the other hand, they should learn English because they live in a community in which the target language is often spoken. The students justify this lack of motivation by saying that they learn English in their classes and always use their native language outside the classroom. To achieve the best results, the course ought to be seen by the learners to be both useful and practical for the performance of their future occupational duties. When the learners have low levels of competence, the intended results may be difficult to achieve. The low standard of achievement in English in Qatari schools is one of the major educational problems that should be investigated to diagnose ways for raising students' motivation to learn English.

1.3.3 Textbooks:

The values and attitudes that are taught at schools are of obvious and central interest to those who are concerned with the social future that the patterns of schooling seem to foreshadow with the view of the world at a certain time. When the values that schools reflect become inconsistent with the values of a certain society, i.e., the beliefs the individuals hold, textbook bias is then obvious. In this way textbooks seems to threaten the individual values. It should be kept in mind that a school subject is always considered a body of information about a field that is thought appropriate for the education of the students. To make the students accept the target information, we should know the social background of the learners.

As for the Crescent English Course, following are some shortcomings of the course.

1-The introduced materials are intended to present the target language in a social context as an outstanding feature of the communicative approach. The fault with these materials is that the target culture is introduced before the native one. To make the the students have the necessary interest in the taught materials, it is better to introduce them to their native culture first. In the Crescent English Course The

Book which is taught in Primary 5, introduces the pupils to a song which runs as follows:

Humpty Dumpty sat on a wall.

Humpty Dumpty had a great fall.

The same pupils will be introduced to their native culture after four years, that is, in Preparatory Two. They will read about Ramadan, which is the month of fasting, Prayers, Zakat, which means alms, neighbours, and Hajj, which means pilgrimage to Makkah. On the other hand, the Crescent English Course has been revised so many times to suit the learners' needs in Qatar, that is, to introduce the native culture first. This can only be applied to the books which are taught at Primary and Preparatory levels.

2)-A common weakness is the writing skill because of the following factors:

a)-The lack of a gradual and clear plan to develop this skill.

b)-A clear trend to concentrate on audio-visual instructions instead of reading and writing. This trend clearly affects the learners' ability to form their own strategy about sentence construction that they need in order to develop the skill of writing.

c)-The existing textbooks tend to encourage the learners to do their homework in groups in their classes. This technique undermines their responsibilities towards

this important task. As a result of this technique, the students do not nurture their individual capabilities in writing.

d)-The textbooks concentrate on oral comprehension questions more than the written form which made the learners lose the chance of sentence construction to develop the skill of writing.

3-The learners always express the lack of right and wrong criteria because of less concentration on grammar, that is, they cannot judge whether a particular sentence is right or wrong. This can be easily seen in answering questions such as "put the verbs in brackets in the right form" .

1.3.4 Tests:

The tests that are constructed either by the Inspectorate or by classroom teachers are still traditional in the sense that they are not analysed to point out weaknesses or strengths in students' achievement. Byrd (1986) in a study about English Language instruction in the Gulf Arab States describes the testing procedures as inadequate and it is generally considered a problem in the teaching-learning process.

> There were a number of reports dealing with students in some states. Grading and evaluation procedures were sometimes lax and failed to represent the real achievement of students. Teachers were reported to be under pressure at times to pass students regardless of their performance. This pressure came from Headmasters/Mistresses, who probably worry about the "image" of their school if there are a lot of low grades. (P. 53)

The teacher's primary role in the classroom is to teach. His success or failure could be ascertained by testing the intellectual growth and development of the pupils' standards. The existing system of testing is not able to verify the pupils' standards of achievement and describe their progress toward the intended objectives.

Tests, in general, aid in determining the learning outcomes of classroom instruction. The teacher-made test is a reflection of what the individual teacher considers important. The teacher can then evaluate success or failure in relation to test results. An analysis of the students' responses on the test can be helpful to the teacher in adjusting the present level and direction of classroom instruction which will affect the standards of achievement in the target language.

In Qatar however, teachers construct classroom tests just to promote students to a new class and nothing else. The decisions about standards of achievement are completely left to the personal intuition of teachers. If they are asked about their students' standards, they have no scientific evidence to

describe these standards of achievement. If teachers of English in Qatar were aware of the role of testing in teaching, they could increase their students' achievement by helping to develop study habits and direct intellectual energy toward the desired objectives.

1.4 THE System oF Testing iN Qatar:

There are two public examinations in Qatar, the first is called the first session exam which is usually held in June, the same exam is called English Promotion Examination. Students who pass this exam in all the seven subjects, Religion, Arabic, English, Maths, Science, Social Studies, and Art Education, will be promoted to the next year of schooling if they get 50 % in all the subjects. Students who fail to get 50 % in three of these subjects, will take another exam at the beginning of the next year, that is in September. This is called a Second Session Examination. Those who fail in four subjects will repeat the same year of schooling again. The following procedure is always followed when teachers intend to calculate the passing mark:

40 % of the total mark is assigned to the ongoing assessment that the teachers carry throughout the school year, two months

before the Mid-year exam, and two months after the Mid-Year exam. (The Total Mark is 40). (Passing Mark is 20).

1st Month + 2nd Month = Total ÷ 2 = Mean 10% is taken.

Mid-Year Exam 20%

3rd Month + 4th Month = Total ÷ 2 = Mean 10%

First Session Exam 60%

10% + 20% + 10% + 60% = 100% [50% pass mark = 20 out of 40]

The English Inspectorate in the Ministry of Education is responsible for constructing the promotion examinations for some stages of learning, that is to say, Primary, Preparatory, and Secondary stages. This situation means that there is a kind of centralization of test construction in Final Promotion Examinations. This centralization passes over teachers' participation and this

creates a detrimental influence on both teaching and testing at the same time because no members of the Inspectorate are qualified in testing and they do not live the real classroom life as teachers do. If the Inspectorate justifies this kind of centralization of test construction,on the basis that teachers lack the adequate knowledge on testing, inspectors can suggest in-service training courses for teachers to practise test construction.

1.4.1 The Testing System Has the Following Scheme:

1st test	2nd	Mid-Year	3rd	4th	Final
Nov.	Dec.	Jan.	Mar.	Apr.	JUN.

Tests that are constructed by the English Inspectorate:

1. Mid-Year examination. (Sometimes).
2. Final examination. (Always).

Tests that are constructed by classroom teachers:

1. 1st month test.
2. 2nd month test.
3. 3rd month test.
4. 4th month test.

Sometimes, the Inspectorate members ask some teachers to construct Mid-Year tests and then the inspectors choose some of these tests.

1.4.2 Examination Times:

Grade	Written	Oral
Primary 5	2 hours	in groups
Primary 6	2 hours	=
Preparatory 1	2 hours	=
Preparatory 2	2 hours	=
Preparatory 3	2 hours	=
Secondary 1	2 hours	=
Secondary 2 (Science)	2 hours	=
Secondary 2 (Arts)	3½ hours	=
Secondary 3 (Science)	2½ hours	=
Secondary 3 (Arts)	3½ hours	=

= All other 1st year Secondary exams-Technical, Commercial, and Religious are as Secondary 1 General.

= All other 2nd and 3rd year Secondary exams-Technical, Commercial, and Religious are as follows:

Written exam : 3 hours. [Oral in groups]

These times are the same for Mid-Year, Fist Session, and Second Session Examinations.

The ongoing everyday activities in classrooms require teachers to use techniques to assess the progress of students. The most often used technique is the administration of a paper and pencil test which presents selected tasks to students. In Qatari schools, students always trust the teacher who can simplify the textbook and make it look easier for them to memorize. I can give an example here to explain what I mean. In Preparatory Three, there is a text about Qatar National Museum which deals with its history. Teachers of English who teach this grade give typical questions together with their answers as a teaching technique.

Q. What is the purpose of Q.N.M. ?

A. To show young people how life was like not very long ago.

Q. What can visitors learn at the museum ?

A. How coffee was made, how tents were put up, and how falcons were trained.

Questions such as these test information not language. As a result of this situation, teachers are sometimes judged by their students on the basis of making the textbook easy to memorize by presenting a variety of ready-made questions and answers. This is because school examinations always concentrate on the quality of information presented by the students. Tests are misunderstood by both teachers and students alike. Teachers still look at testing as an unbearable responsibility which is imposed on them by educational accountability. Testing the target language on this basis kills the creative ability of the students and encourages the ability of memorization which may be used as a way of learning but not as a method of testing. In this type of testing, it is difficult to determine the students' communicative ability.

Teachers in the Qatari schools use different kinds of test to measure their students achievement. The following notes will explain the situation.

1-At the primary level, all the test instructions are in the native language. The justification for that is to ensure that no pupil fails to answer a question because he cannot understand the instructions, teachers always explain the main test format. The type of questions that are used at this stage

swing between the completion items and description of people. Vocabulary tests are always used in isolation. Grammatical items may sometimes be used to ensure that the pupils master "Verb to be" for example. The use of illustrative pictures is often part of any given test. These pictures are intended to test "Numbers" or parts of the body.

2-At the preparatory level, all test instructions are also given in the native language of the pupils. The same justification as at the primary level can be applied here. The types of test are either objective or subjective. Some tests may have both objective and subjective items. Test instructions usually begin with an imperative e.g. [Write a reply.....], [Use these notes to write .....], [Answer these questions.....], [Change these sentences as shown in brackets.....], etc.

3-At the Secondary level, all the test instructions are in English. The English Inspectorate in the Ministry of Education distributed a handout as guidelines for teachers to help them in their task when they want to construct classroom tests. The following notes were included in the handout:

1- Writing From Notes:

This type of question appears in several forms: "Use these notes to write a report / reply / describe....." etc.

Pupils should first study the notes and see how they relate to the subject to be written about.

2- Direct Questions:

"Read the following passage and answer the questions below." This sounds simple, but few students know how to tackle such a question in the most efficient manner.

3- Transcoding:

This type of question depends on the pupil's ability to read charts, diagrams, maps, etc. and transfer the information into sentences or vice versa.

4- Direct and Reported speech:

Two types of activity are possible here:

a) a scene is described and the pupil has to write a dialogue of what might have been said.

b) after reading a dialogue pupils have to report what were said e.g. " When I interviewed the prisoner he told me that....."

5- Cloze Test:

This test is used in various forms:

a) Filling in the blanks with a word appropriate to the context.

b) Choosing the word or phrase which means the same as the underlined word.

c) Putting a verb into its correct form.

6-Matching sentences:

Pupils are asked to read sentences and match them with pictures or two groups A and B and the students are asked to match the sentences that go together, most of these sentences are taken directly from textbooks.

7- Translation: (Secondary two and three Arts only)

A short passage of 6-10 lines to be translated from English into Arabic.

8- Topic sentences: (Secondary Stage only)

Pupils are asked to identify and copy the topic sentence of a short paragraph. They must be able to distinguish

the sentence which states the main purpose of the speaker.

As for oral testing in the Qatari school, it is still carried out traditionally. What I mean here is that testing students' ability to use the target language is not practiced as it should be. At all the three levels (Primary, Preparatory and Secondary) teachers usually ask the same set of questions. Some of these questions may include:

-What is your name?
-How many brothers and sisters have you got?
-Where do you live?
-How did you come to school?
-Tell something about your country?
-What is the man doing in this picture?

These questions and others are often asked on an oral test. Some teachers get bored when they hear the word oral. They forget the fact that oral testing is one opportunity for students to apply what they have learned by expressing their ideas and points of view with previous preparation in the foreign language. This activity presents a difficult problem for the teacher when it comes to evaluating student performance in such a complex combination of skills. An oral presentation entails not only grammar and pronunciation but also the student's communicative ability and the organization of his

ideas. When we come to the rating scale in the present situation, teachers do it itintuitively. Most students pass the oral test, not because most students are well qualified in English, but because of the absence of criteria.

In a word, testing in Qatar is in need of urgent modification. It is still carried out traditionally because it devotes no attention to such matters as reliability and validity.

1.5. Problems of Testing:

Constant attempts have been made by inspectors and teachers to improve school examinations, but within the confines of the traditional pattern. A technical unit or section of educational evaluation is not yet established in Qatar. We hope that the day will soon come when such a unit will be established. In 1990, the Ministry of Education has appointed Mr. Roger Nunn, native speaker, as a testing co-ordinator for only one year. Mr. Roger Nunn has been transferred to the English Language Teaching Unit (ELTU) at Qatar University to teach English, but the evaluation section is still a dream. There are many problems that undermine testing in Qatar, among them are the following:

1. Old Techniques of Testing:

Although Qatar has adopted contemporary materials, testing procedures are still traditional. Tests have to test what is taught, but teachers teach communicatively and test their students traditionally.

2. Lack of minimum knowledge on testing and its role in teaching. Teachers test their students just to promote them to a higher stage of learning. They often use personally invented techniques that are repeatedly applied in the classroom. The success of up-to-date English language teaching approaches requires the knowledge of specialized expertise to train and provide teachers with suitable and recent procedures of testing.
3. Present techniques do not make a reasonable balance to emphasise the four skills. Although the communicative approach concentrates on the four skills while teaching, testing pays special attention to writing (ticking off or circling the correct answer, filling in blanks, answering questions). The current testing procedures do not treat oral or listening skills in the real meaning of testing, and fail to focus on reading skills. Testing what has been taught is a failure.

4. The testing methodology is at odds with classroom teaching procedures. There is no relationship between testing and teaching. Testing should be looked at as an indicator of the students' achievement in the classroom.
5. Ignorance of the teachers' role in the present process makes them negative in improving their knowledge on testing. To get the best results, teachers should be familiarized with various techniques of testing through in-service training for both newly appointed teachers, who lack the appropriate methodology of testing, and experienced teachers who are also in need of such training courses to learn how to test communicatively.
6. Little attention has been paid to the importance of testing in improving teaching.
7. No firm indication is given to the importance of reliability and validity of the tests being used in the classroom.

Testing English as a foreign language in Qatar plays an important role only in deciding the students' future position, tests are not analysed to point out strengths and weaknesses, so testing can be looked at as an objective in itself more than a means of improving teaching. Once the results of the examinations appear, the role of testing is finished. In the

present situation, there is no evaluation of test scores that should be done by the classroom teacher in order to diagnose certain weaknesses and strengths of the students' achievement. If some of these tests are analysed by the Inspectorate, teachers are not informed of the results of this analysis.

The teachers' role cannot be easily ignored since they are the only people who live moment-by-moment with classroom activities. They know where exactly their students are. This enables them to diagnose their students' weaknesses. The Inspectorate's justification for the centralization of testing is that teachers lack the necessary knowledge of test construction and interpretation, although none of them is well qualified in the testing domain either. It is however true that teachers lack the necessary knowledge about test construction, but we should not leave this ignorance to continue. Teachers should know how to describe their students' achievement, how to choose test items and how to assign test scores.

In order to describe in a little detail the lack of knowledge among teachers, I shall give a brief account of some classroom tests. They are teacher-made tests, although the Inspectorate sets sample tests that could be followed by classroom teachers when they want, either to train their pupils on such items, or to imitate these tests:

1-A short passage which is usually followed by four or

five comprehension questions. (it is sometimes taken from the textbook).

2-One essay question. "Write a letter to your pen-friend, tell something about your country", or "Describe this animal"

3-A cloze passage of about 4-6 randomly selected gaps.

If we consider these three examples, we will find that two of them place great emphasis on reading and less on writing and they place the testee in a position to recognize rather than to create responses. As a consequence, examinations in Qatar are still traditionally constructed in such a way that teachers with three or four years of experience can easily predict most test items to be included in any test. Items are just stereotypes repeated now and then.

What is happening in Qatar now is that because teachers test their students just to pass them to a higher stage of learning and nothing else, test item construction is always a fast procedure. When a teacher wants to construct any test, he does not spend the necessary time on test item construction, and classroom teachers do not care about the necessary qualities of a good test because they do not know how to estimate either validity or reliability. Both are effective procedures to obtain effective testing and should be understood by any one working on testing. Validity, content validity in particular, is very necessary to be understood by classroom

teachers to make sure that they really test what they have taught to their students and nothing else.

Since the communicative approach is used in teaching in the Qatari schools, tests should be constructed communicatively to suit the activities the approach is trying to provide. Communicative testing as seen by classroom teachers is no more than a test which pays no attention to the accuracy of the students' responses while scoring the test. As a result of this belief, they begin to accept any response even if the response is incorrectly written. Teachers justify this on the basis that the communicative approach does not pay attention to the language structure but it stresses the idea of communication and language function.

Another factor in the situation is that teachers have always been asked by the Inspectorate to train their students on certain kinds of questions such as the following: (Prep. Three Level).

Do as shown in brackets and rewrite the sentences again.

==========

1-A man (Drive) his car now. [Correct the verb]

2-a.s.a.p. [Write in full]

3-Why do a lot of fish die in rivers ? [Answer]

4-A woman who keeps medicine at home. [Give warning]

5-Man is destroying his environment. [Give two examples]

When the students are asked to answer such questions in the mid-term exam, they all begin to ask teachers of English about how to answer this particular question (Do as shown in brackets). I think the students cannot answer the question for two reasons. The first is that they cannot understand what is meant by the words in brackets, and the second one is that their teachers do not train them in how to answer this kind of question. They have not been introduced to such questions before.

1.6 The Need and Origin of the Study:

As a consequence of the situation described above, I felt that there is a lack of knowledge on testing and its role in education among classroom teachers. As a result of this feeling, I decided to concentrate on testing to improve my views on the target domain on the one hand, and those of my colleagues on the other. I felt the need for this study during my fifteen years of teaching English in Preparatory and Secondary schools in Syria and Qatar. If teachers are asked about the kind of test they use when they test their students, they cannot classify the tests they often use as achievement,

or diagnostic as criterion-referenced or norm-referenced tests.

The present system of testing needs quick modification to cope with the new techniques of teaching, the shift from the Structural to the Communicative approach in teaching EFL. To achieve the best results, teachers should acquaint themselves with the most up-to-date testing procedures. Teachers of English in Qatar are in need of adequate knowledge about the following topics:

1. Test function.

   When we test our students in the Qatari schools, we do not know exactly why we test. Do we test to promote our students to a higher stage of learning or to measure what the students have achieved ? It should be pointed out here that testing has to measure positive and negative achievement. It seems to me that we test just for promotion as far as the Qatari schools are concerned.

2. Test format.

   The school test should be a reflection of what the students have studied during a limited period of time.

3. Test effect.

Tests in general have advantages and disadvantages on the teaching-learning process. For example, when the students are tested orally, the classroom teacher will concentrate on preparing the students for the oral performance.

4. Teacher's role.

Teachers should be given further consideration. This means, teachers should be given more freedom to prepare their tests because they live the classroom life in its real meaning. Those who direct the teaching-learning process, are not able to diagnose the students' problems either in teaching or in testing. Hence, the need for this study is of great importance to classroom teachers to improve their ability to test correctly and efficiently.

Chapter Two:

OVERVIEW OF HISTORY AND DEVELOPMENT OF LANGUAGE TESTING:

Topical Outline

2.1 Introduction

2.2 Overview of history and development of language testing.

2.2.1 The early intuitive stage (before 1850)

2.2.2 The pre-Scientific Stage. (1850-1950/60)

2.2.3 From 1900 to 1915.

2.2.4 from 1915 to 1930.

2.2.5 From 1930 to 1945.

2.2.6 From 1945 to 1960.

2.2.7 Modern testing

2.2.8 Testing in the 1980s

2.3.1 The historical development of testing in Qatar

Chapter Two

2.1 Introduction:

In chapter One I provided a discussion of the main issues and problems of testing in Qatar. In this chapter, I shall deal with the stages of development in language testing in order to locate the Qatari situation and begin to determine the stage we have reached in Qatar. After a review of the main periods in the history of testing, I shall deal, in more detailed perspective with the current state of testing, that is to say, the progress that has been made in testing from 1980 to 1988.

2.2 Overview of History and Development of Language Testing:

There is a problem one faces when trying to overview the the literature about early stages of testing English as a foreign language. This problem is mostly related to the difficulty of investigating different views of different people and making them come near one another to establish a kind of authentic and harmonious division.

The former history of language testing can be divided into two major periods of development according to Madden (1983). The early intuitive stage, and the scientific stage,

> Testing during the last century and the early decades of this one was basically intuitive...After the intuitive stage, testing entered a scientific stage. (PA. 5-6).

2.2.1 The Early Intuitive Stage (before 1850)

In this stage testing depended on the personal judgment of the examiner and all examinations were performed orally. Since testing was oral and it did not depend on well defined types, the scores were variable. Thorndike and Hagen (1969) explain what was happening in this stage:

> Before 1850, testing had relied very largely upon oral examination. The teacher or visiting examiner asked a question. The designated pupil undertook to answer it. The questioner arrived at an immediate subjective evaluation of the answer. The method was burdensome and inefficient since only one pupil could be tested at a time and since different pupils were asked different questions, the answers to which were evaluated subjectively by the examiner (P. 2).

It is clear from the previous quotation that testing was basically dependent on the testers' personal judgment simply because testers were untrained. As a result, both teachers and testers tried their best to invent ways of teaching and testing, but unfortunately, these invented ways concentrated on language structure instead of language use. Madsen (1983) stresses the point when he says:

> Facts about English often weighed as heavily as skill in using the language. As a result, students had to label parts of a sentence and memorize lists of language pattern (I am, we are, you are, she is, he is, etc. (P. 6)

The early intuitive stage lasted for a long time during which there was a reliance on the knowledge of grammar taught in a traditional method. This stage of oral testing was criticized because more time was needed to test a group of pupils, and students were asked different questions. The aim of testing had not been achieved and no comparability was then possible among pupils. The testees were not offered the same opportunity either in the task or in the mode of evaluation. The next stage of development began to overcome these disadvantages by turning to written tests.

Another division of the target domain was also classified by Spolsky. Madsen (1983) quotes Spolsky (1978) who says:

> In 1978, Spolsky identified three major historical trends in language testing, which he referred to as "pre-scientific", "psychometric-structuralist" and "integrative-sociolinguistic" (P.432)

2.2.2 The Pre-Scientific Stage. (1850-1950/60)

In this period, oral tests were replaced by written examinations as a basis of promotion or even admission to colleges. The technique which was used most widely in this

period was the essay. Thorndike and Hagen (1969) explain the justifications for this change from oral to written examinations:

> The written examinations had advantages over the oral examinations of (1) presenting the same task to each examiner of the group, (2) letting each pupil work for the full examination period. (P. 2)

During this stage of testing, little attention was paid to reliability and validity. Testing was carried out in the form of translation and reading approaches, and exercises on selected grammar points dealing with general rules of language.

In the second half of the pre-scientific stage, subjective written tests began to be replaced by objective tests because the latter could be scored consistently even by untrained people. Specialists started to evaluate tests statistically, looking at the effectiveness of each item. This led to a new consideration of reliability and validity of tests.

It was believed that the second period, the " Psychometric-structuralist " was developed by Lado. To act with confidence, it is relevant to quote Madsen (1983) who says:

> The " Psychometric-structuralist ", may be considered to have evolved in very large part from the theoretical work and related empirical studies carried out by Lado. (P.431)

This trend in testing suggests that when trying to test students' ability in the target language, students should be exposed to a variety of problems during the test. These language problems should be limited to the same points of difficulty. That is to say, all the presented questions should be dealt with at the same level of difficulty. The Psychometric-structuralist can be looked at as an advanced level of testing because it pays special attention to the students' ability to solve problems that are related to the target language. If they are able to deal effectively with these problems, they can be considered as masters of the target language. Madsen (1983) quotes Lado (1957) to explain what is meant by language problems and their level of difficulty.

> Since some , aspects of the target language, are easy to master they are already from previous language training in mastering the native language, we will generally eliminate these from the corpus. We will attempt to test the learning problems, on the ground that knowing the problems is knowing the language. We say specifically that testing the problems is testing the language. (P.432)

In this period a new type of testing had emerged and was completely dependent on linguistic theories. That is, the

discrete-point approach in teaching and testing which assumes that the target language can be broken into separate elements and skills and can also be taught and tested on this basis to diagnose the learners' ability to use the target language. Oller (1979) sums up the aims of this approach:

> These three goals, that is, diagnosing learner strengths and weaknesses, prescribing curriculum aimed at particular skills, and developing specific teaching strategies to help learners overcome particular weaknesses, are among the laudable aims of discrete point testing. (P. 211)

But this way of testing was criticized on the basis that we harm the language stream when we isolate it into pieces and put it together again. Oller explains the problem of this ineffective separation.

> Discrete point analysis necessarily breaks the elements of language apart and tries to teach (or test) them separately with little or no attention to the way those elements interact in a larger context of communication. What makes it ineffective as a basis for teaching or testing languages is that crucial properties of languages are lost when its elements are separated. (P. 212)

The third period in Madsen's division is the "integrative-sociolinguistic" which assumes that the learners' ability can be tested on a wider basis than discrete point approach does. This period showed a clear rejection of the

discrete-point approach on the basis that the use of language for real-life communication involved a creative act in which the whole of the communicative event was considerably greater than the sum of its linguistic elements. As a result, the adequacy or effectiveness of the communication could not be adequately assessed through individual evaluation of its component parts.

The historical development of twentieth century testing indicates new intentions to test students objectively. So ways were needed to provide objective information to get a sound judgement of students' progress and of a school program's effectiveness as well. The first 60 years of the twentieth century are divided by Thorndike and Hagen (1976) into four equal parts:

2.2.3 From 1900 to 1915.

This was a period of exploration and initial development of methods. It saw the emergence of the first Binet intelligence scales. Standardized achievement tests in different subjects began to appear. The appearance of achievement tests led to a systematic way of testing.

Since new ways of doing things are usually contrived because of a pressing need, achievement testing came into being as a need for knowing students' achievement in a particular and

practical school situation. During this period, standardized tests came into existence as a direct result of the early efforts of those who were looking for a uniform method of measuring children's abilities; intelligence testing also contributed to this period when Binet gained wide recognition of his work as a psychologist. Tuckman (1975) says:

> By 1904 Alfred Binet had established himself as France's premier psychologist and expert in human individual differences with his studies of the differences between "bright" and "dull" children. (P. 15)

2.2.4 From 1915 to 1930.

This period can perhaps be called the "boom" period in which the result of achievement testing received much attention, but less attention was given to the developments in writing test items, although test items are the basic building blocks of tests. Pioneers in testing had shown their ideas to their enthusiastic followers, tests were multiplied. Standardized tests were developed for all school skills and for the content areas of school programs.

Test results were unquestioned. Thorndike and Hagen (1969) explains this trend in testing:

> test results were often accepted unhesitatingly and uncritically and served as the basis for a variety of

> frequently unjustified judgments and actions with respect to individuals. (P. 6)

2.2.5 From 1930 to 1945.

Thorndike and Hagen give a short description of the period when they say:

> From 1930 to 1945 may be considered a period of critical appraisal ... It was a period in which the centre of attention shifted from measuring a limited range of academic skills to evaluating achievement of the whole range of educational objectives (P. 6)

2.2.6 From 1945. to 1960

During the present century, many testing techniques have been developed and characterized by many educational changes. There was an urgent need for new techniques in testing to go hand in hand with new methods in teaching on the one hand and to help classroom teachers to understand the new terminology used in reports of ongoing testing of students' achievement on the other hand. Valette (1977) says:

> It is clear that these new methods of teaching require new methods of evaluation. (P.7)

Thorndike and Hagen (1969) describe this period as a good opportunity for standardized testing:

> The mid-twentieth century is a period in which standardized testing is a widely experienced and widely accepted phenomenon. (P. 7)

2.2.7 MODERN TESTING

The late 1960s and 1970s were clearly a period where assessment of students became systematic. Many new books were written about testing and assessment. Doyle (1984) mentions a long list of writers, among them are the following:

> Bottom (1973), Miller (1972 - 1974), Page (1974), Doyle (1975) and more recently, Grasha (1977), Centra (1979) and Millan (1981) ... All these books indicate the evolution of instructional evaluation. (P. 5)

But these tests failed to measure outcomes of high school programs. Then there was a marked shift from the linguistic to the communicative dimension. The emphasis was no longer on the perfect linguistic accuracy of the learner's responses, but on the ability to function effectively when using the language in real situations. At this stage, the most important change was the emphasis on the learner's knowledge and use of the target language. communicative testing began to take its form as an adequate alternative to tests that concentrate on the accuracy of the learner's ability to use the

language he learns.

In this period, that is in the 1970s, little knowledge was available about the nature of the communicative proficiency of the learner. In an article in 1979 Morrow began by quoting Wilkins (1976) to express the difficulty in dealing with communicative language testing:

> Wilkins (1976) concludes with the observation that, "we do not know how to establish the communicative proficiency of the learner". (P. 9)

In summary, we can say that language testing has been developed enormously in recent years and has also absorbed many influences. We no longer believe in the most common characteristic of testing in the past namely that testing has nothing to do with the teaching-learning process. We also no longer believe that there is a single kind of language testing, since different kinds of tests are needed for different situations. The importance of education has gained wider public appreciation. Teachers, administrators and the general public need accurate information about educational progress to reach decisions about further development. This will be the main topic of the following section.

2.2.8 TESTING in the 1980s:

This section will be devoted to the characteristics of testing in the 1980s. This review of the current state of testing will be an attempt to outline the progress that has been made in testing from 1980 to 1988. Such a review will be useful for establishing the base line of our own approach. It will also throw light on current procedures.

In recent years, testing and assessment scholars have learned a great amount about testing. Testing has become much more complex than a simple technical issue and they have invented many new approaches to testing to suit the current approaches of teaching, as a result of this, interest in testing has grown.

The purpose of assessment now goes beyond the analysis and evaluation of a particular case or program. Its scope has been expanded to include how to make the best use of these tests to employ the results in improving the teaching-learning process.

It is difficult for any researcher to trace all the recent changes and developments that have taken place in testing. Many changes and new developments seem to be merely fashion, that is, they quickly rise and more quickly fall. Other changes and developments have become well known and accepted as a foundation for many other developments.

The current state of testing presents greater emphasis on

the use of tests to improve learning and instruction. Tests are no longer dependent solely on paper and pencil, since there are a number of important changes that require the use of new procedures. To improve the present state of education is a challenge. It needs to resist powerful traditions and influences. Many of the aspects of achievement that teachers wish to assess in the present time cannot be assessed by the traditional techniques, so the need for alternatives is evident. Now, testing techniques go hand in hand with teaching procedure.

2.2.8.1 The first characteristic of current testing represents an increased use of criterion-referenced testing which has the ability to tell us about a student's situation. Brown (1981) says:

> Criterion-referenced assessment compares the pupil's attainment with a criterion. It identifies what a pupil knows or has attained. (P. 2)

Criterion-referenced testing tells us about the knowledge of the subject or performance of skills that is of concern. The concerns that have led to the establishment of the increased use of criterion-referenced testing include the need to know what students know or can do when using the target language.

2.2.8.2 The second characteristic is graded tests that have frequently arisen because of dissatisfaction with traditional methods of both teaching and testing. Murphy and Torrance (1988) quote Pennycuick (1986) who explains what is meant by graded test.

> In a graded test scheme, there is a sequence of tests at progressive levels of difficulty, complexity, sophistication and or syllabus content, which are designed to be taken by students only when they have a high probability of success. Each test is closely linked to the curriculum for the relevant level by means of clear specification of the knowledge and processes to be assessed and of the standards to be attained. (P. 68)

Murphy and Torrance also consider these tests as a new area to be discovered. They say:

> The graded test movement is still in its infancy, there is still a great deal to be learnt. (P. 2)

The most interesting developments and also actual progress in graded testing have been teacher-led. In Britain, the clearest example of this is the Graded Objectives Movement in Foreign Language Teaching. Skehan (1988) gives justifications for this movement:

> Dissatisfied with traditional examinations, and also more concerned with communicatively oriented teaching, graded objectives schemes have been

> developed in many parts of the country (Britain). (P. 220).

Graded tests are seen as high-quality teacher-made tests for internal assessment of students and as an alternative to external examinations. Those who praise graded tests are far more concerned with the effects of these tests on classroom practice than they are with technical aspects of the tests. Murphy and Torrance evaluate the state of the graded tests saying:

> A particular issue which remains unsolved is whether schemes are best designed for a wide or restricted ability range. The long-term motivational effects of graded tests are not clear, and there are technical problems (e.g. in the description of pupil achievement) which may continue to be troublesome. However, graded tests represent a significant component of the recent surge of assessment initiatives, and deserve further evaluation. (P. 84)

2.2.8.3 The third characteristic of modern testing is formative testing which feeds into the process of learning and provides teachers and students with information that may change or at least modify behaviour while teaching is still relatively rare. These trends have increased confidence in teachers' accountability for improving learning and instruction, and imposed a great responsibility on them, especially in the U.S.A. Gronlund (1981) says:

> In recent years, a number of states have passed laws making teachers and other school personnel accountable for the learning and development of students. (P. 13).

To cope with these demands, teachers need their tests to be as consistent as possible. To achieve this quality, teachers must pay attention to reliability to know how consistent scores are, and this will be discussed in a later chapter.

2.2.8.4 The fourth characteristic is the use of the computer which has had a profound impact on the rapid development of testing. Testers expect that the use of the computer will have greater influence in the near future. Madsen (1983) emphasizes this saying:

> A current trend in the language teaching field- a trend that has not yet been explicitly related to language testing to any applicable extent but that shows considerable potential for becoming an important and integral component of the overall measurement operation is the use of computers in the service of language instruction. (P. 436)

The most apparent contribution has been seen in the scoring and analysis of tests. The type of feedback made possible by using the computer has been a major factor in improving both learning and instruction. When school computers are widely used, teachers will be able to prepare classroom tests of high quality and avoid much of the routine

works to construct their tests.

In summary, we can say that research in language testing has made important strides recently, but significant problems remained unsolved. Although language testing has always had its friendly and unfriendly critics, in recent years there has been increasing concern about the role of testing in the school One criticism is that most of the testing procedures currently used to assess students or school programs cover a narrow range of the knowledge and skills that are commonly considered as the goals of schooling.

It is often argued that the face of language testing has already changed and most testers think that it is likely to continue to change for some years to come. In the future, clearly teacher involvement in assessing the learning process will be a prerequisite for the realization of the expected ideas and practices. This involvement seems likely to continue and will maximize the opportunity for teachers to experiment with different sorts of course work assignments and gain experience on improving the quality and effectiveness of their teaching methods and school syllabus.

2.3.1 The Historical Development of Testing in Qatar

In chapter one, we provided a detailed discussion about the factors that affect students' standard of achievement, and

among them we mentioned classroom tests. These tests are constructed traditionally just to promote students to a new position. They have "little to do" with students' achievement during the school-year instruction.

Testing in Qatar has undergone major changes since the introduction of the Crescent English Course. The Ministry has appointed a full-time testing co-ordinator who has attended a British Council course on testing communicative competence. He used to write all the end-of-year examinations, both public and promotion; the mid-year examinations are sometimes written by classroom teachers. At the beginning of the application of the English Crescent Course, all levels from Primary Five to Secondary Three had a half hour listening test (the instructions were in Arabic) and a fifteen minutes oral test. The oral test performance was different in accordance with the level of the students. At the Primary level, the oral test was conducted by the classroom teacher. At the Preparatory level, by the classroom teacher; while at the Secondary level, the oral test was conducted by the Inspectors and English Language Teaching Development Centre (ELTDC) staff paired with Secondary teachers. Students were examined in groups of four. Before the use of the Crescent English Course, when the Audio-Lingual Approach was in use, there was neither a listening nor an oral examinations.

Through the stages of test development in Qatar, tests began to integrate writing and reading and an attempt has been made to produce real-life tasks. Tests began also to avoid discrete-point multiple choice testing and questions like: "Read the following and answer the questions" and "Write a short paragraph about your last summer holiday". To increase the teachers' ability to construct classroom tests, the appointed testing co-ordinator tried to establish a test-bank of items covering all the levels of teaching (Primary, Preparatory, and Secondary). These items were available to teachers as models to be followed when constructing in-class tests and sometimes in mid-year examinations.

In 1982, Brendan Carroll, testing expert, came to Doha in a series of consultancy visits at the request of the Ministry of Education. These consultancy visits were arranged by the British Council in Doha. In the final report, Brendan Carroll states the purposes of these visits to be as follows:

1. To comment on the effectiveness of the present approach to English Language testing and examining in Qatar schools.
2. To conduct workshops and discussions in the construction of tests and examinations of English in the context of the ideals of the Crescent Course.
3. To make initial steps in an overall objective

> assessment of the progress of children, especially those at the upper end of the school system, in their mastery of English as a means of communication. (Carroll: 1982 -P.1)

Constant attempts have been made to test communicatively and an impressive amount of work has been done on devising tests that reflect the aims of the Crescent Course. As a result of these attempts, teachers were asked to carry out an on-going assessment of their students. The idea is to urge teachers to note the progress of each individual in the class to take the appropriate remedial action, particularly with those who fail the tests.

The present state of testing, in Qatar, places great emphasis on testing school students just to move them to a higher level of education. Tests in the Qatari schools are not used in the real sense of testing. They have nothing to do with teaching improvement. There are many suggestions to improve the current state of testing in Qatar, but they are all still theoretical. Although many attempts have been made to improve the current state of testing, little knowledge has been learnt about classroom testing. We are in need of a " Testing Unit " to provide us, as teachers, with the most up-to-date procedures and techniques of testing that enable us to use test results to improve teaching and direct it in the desired way. Teachers

teach communicatively and still test traditionally. They do not care much about whether the course objectives have been achieved or not, because test results mean nothing to classroom teachers except students' promotion

The evident theoretical period in the Qatari system of testing is the modern period where the communicative approach is of central interest , but no distinct signs of communicative testing are being applied in the present situation. The present state of testing is no more than miscellaneous procedures that escape clear-cut classification.

In Qatar, teachers of English are in a vantage position to make every possible effort to benefit from the communicative approach currently in use in the Qatari schools to test communicatively, but the current procedures are always faced with administrative problems. Qotbah (1990) deals with the matter irrefutably:

> Although the communicative test has its own approach to language testing which makes it different from other approaches, it is part of the educational system and can be affected to a great extent by administrative problems. In other words, the students' percentage success in exams is a criterion assigned by the Ministry of Education to gauge the academic reputation of schools. (P. 41)

With regard to the Qatari testing state, the time is highly suitable to dismiss the traditional techniques of testing and to apply the modern techniques in language testing.

Involvement of teachers in the testing process should be more obvious than before. We mean that teachers should test their students and deal with test scores as a good indicator to do what should be done to improve the current state of teaching.

Chapter Three

PURPOSES AND KINDS OF TESTS

Topical Outline

3.0	Introduction
3.1	The significance of testing in the classroom.
3.2	Purposes of testing in the classroom.
3.2.1	Measuring outcomes of instruction.
3.2.2	Improving the curriculum.
3.2.3	Placement and grouping.
3.2.4	Predicting future performance.
3.2.5	Describing realistic standards of achievement.
3.2.6	As a diagnostic procedure.
3.2.7	To grade students for promotion.
3.3.1	Different types of test.
3.3.2	Achievement testing.
3.3.3	Standardized tests.
3.3.4	Teacher-made tests.
3.3.4.1	Advantages of teacher-made tests.
3.3.4.2	Disadvantages of teacher-made tests.
3.3.5	Norm-referenced testing.
3.3.5.1	Advantages of norm-referenced tests.

3.3.5.2 Disadvantages of norm-referenced tests.

3.3.6 Criterion-referenced testing.

3.3.6.1 When to use criterion-referenced tests.

3.3.6.2 Criterion-referenced test construction.

3.3.6.3 Tasks for level 1. (lower).

3.3.6.4 Validity and reliability of CR tests.

3.3.6.5 Advantages of criterion-referenced tests.

3.3.6.6 Disadvantages of criterion-referenced tests.

3.3.7.1 Differences between CR tests and NR tests.

3.3.8.1 Objective testing.

3.3.8.2 Advantages of objective testing.

3.3.8.3 Disadvantages of objective testing.

3.3.8.4 Types of objective testing.

3.3.8.4.1 True-false items.

3.3.8.4.2 Advantages of true-false items.

3.3.9.1 Matching items.

3.3.9.2 Advantages of matching items.

3.3.9.3 Disadvantages of matching items.

3.3.10.1 Multiple-choice items.

3.3.10.2 Advantages of multiple-choice items.

3.3.10.3 Disadvantages of multiple choice-items.

3.3.11.1 Subjective testing.

3.3.11.2 Essay test.

3.3.11.3 Scoring the essay test.

3.3.11.4 Advantages of the essay test

3.3.11.5 Disadvantages of essay test.

Chapter 3. PURPOSES AND KINDS OF TESTS.

3.0 Introduction.

In chapter two I characterized the historical trends up to the present day, and provided a review of the history and development of language testing. My purpose in this chapter is to review the purposes and kinds of test used by teachers in classrooms in general, and by teachers of English in Qatar, in particular. Essentially this will mean concentrating on different purposes of tests as well as on achievement tests in particular and contrasting them with other tests. I shall state the kinds of testing procedures that will be discussed in the thesis. My main concern is classroom testing that enables us, as teachers, to obtain valid, reliable and useful information concerning our students' performance. I will then consider the issue of criterion and norm referencing for achievement tests. I will consider different techniques of test construction. Finally, I will mention what kinds of tests, referencing and construction techniques are used in Qatar.

3.1 The Significance of Testing in the Classroom:

Classroom tests, in particular, play many roles in teaching, among them are the following:

1-Determining progress towards course objectives.
2-Stimulating students' learning.
3-Evaluating classroom attainment.
4-Diagnosing problems.

Through frequent testing, teachers can determine which aspects of the target program are presenting difficulty for individuals or for the class as a whole. When analysing the scores, the teacher can decide where the problems are. Such analysis can also enable the teacher to diagnose the mistakes made on a given test to decide where exactly the remedial teaching should concentrate. Tests enable classroom teachers to discover whether the class objectives are being met or not. On the other hand, teachers can evaluate the effectiveness of a new method being applied in teaching.

The idea of integrating testing with teaching is not new. The topic was presented as early as 1939 by Purnell and Davis when they wrote about using classroom tests to direct learning. On the other hand, Hopkins and Antes (1985) mention what Tyler had pointed out in 1951:

> Educational measurement can have a profound influence in the improvement of instruction, but to do so, it must be viewed as an integral part of instruction, its planning must go hand in hand with instructional planning, and the result must be used continuously to guide the planning and development of curriculum.(P.30)

The purpose of classroom testing is to allow the teacher to meet specific learning needs. After completion of part of the learning course, remediation allows teachers to overcome deficiencies pointed up in the testing procedure that has been applied in the classroom. The frequency of tests during instruction are determined by how instruction is structured. At times, a test may be needed for one lesson and at other times a test may cover a unit. Ideally tests would be given at natural breaks in the instructional sequence. Teachers should space the time between classroom tests to give the needed time for remediation. Testing is commonly considered as a part of every teacher's day. Teachers who do not test to evaluate both their own and their students' work cannot do their job properly, Satterly (1981) emphasizes the idea saying:

> Assessment of a wide range of pupil performance is an integral part in monitoring the progress made not only for the benefit of the pupil and teacher, but those outside the classroom such as other teachers,other schools. Employers will depend upon reliable knowledge of pupils' achievement.(P.1).

The concept of testing has implications for almost every aspect of teaching. Decisions that teachers make about everyday activity help them to predict the students' future performance. Teachers need to know whether the classroom objectives are being met or not. They cannot leave such matters to chance. Without a clear set of goals, the determination is quite impossible. When we say that our students are progressing in the right direction, we mean that we have measured their abilities towards specific goals to reach such a decision. Gronlund (1981) says:

> The more accurately we judge our pupils, the more effective we shall be in directing their learning. An understanding of the principles and procedures of evaluation, then, should aid us in making more intelligent decisions in directing pupil progress toward worthwhile educational outcomes.(P.4).

Madsen (1983) gives further reasons for testing saying:

> Tests can foster learning by their diagnostic characteristics, they confirm what each person has mastered and they point out those language items needing further attention.(P.4).

It is unfortunate that many tests in the past have led to a separation of testing from teaching. Both testing and teaching are so closely interrelated that it is impossible to work on either domain without working on the other. The

essential purpose of tests is to reinforce learning and motivate students.

Teachers used to think that the aim of testing is to point out students' ignorance or lack of knowledge in a certain field of study, but recent studies on testing have emphasised the opposite. The previous view of testing harms the teaching process and restricts students' motivation.

In the following section I shall consider in more detail the main purposes of classroom testing. Finocchiaro and Bonomo(1973) explain the modern role of testing in classroom teaching:

> The three principal reasons for testing as far as classroom teachers are concerned are:a) to diagnose the specific features of language in which individual students or groups are having difficulties; b) to help us guage our ability as teachers ;c) to find out how much our students have learned or achieved (P.206).

3.2. Purposes of Testing in the Classroom:

The fundamental purpose of classroom testing is to provide information for making decisions, that is, for evaluation. Evaluation comprises essentially two components: (1) information and (2) judgments, or decisions. The various purposes of testing can be best understood by dealing with them separately.

3.2.1 Measuring Outcomes of Instruction:

Tests help teachers in deciding the learning outcomes of classroom instructions, because the teacher can then evaluate the success and failure of classroom learning in relation to test results. If the teacher wishes to adjust the level and direction of classroom instruction, he can easily do that by analysing students' responses to the test. The purpose of classroom tests is to help students achieve a set of learning outcomes. These outcomes are established by the instructional objectives, the desired changes in students are brought about by the planned learning activities, and the students' learning progress is periodically evaluated by tests and other evaluation devices.

The first step in both teaching and testing is that of determining the learning outcomes to be expected from classroom instruction. The learning outcomes can also provide guidance about the language skills and functions to be included in tests used to assess students progress. The procedure of stating specific tasks, then teaching and testing them is used in programmed instruction. This makes it possible to analyse each intended learning outcome in considerable detail and to describe the expected student performance at every specific level.

In Qatar, testing did not achieve its goals in measuring

the outcomes of teaching in a communicative atmosphere because the present techniques of testing do not pay the necessary attention to analysing the students' responses to a particular test.

3.2.2 Improving the Curriculum:

Tests in schools should be purposefully related to what schools are trying to accomplish in term of curriculum development. If the purpose is to evaluate the curriculum and develop it, testing should be adopted during the early stages when new materials are being tried. Evaluation data enable the curriculum developer to determine the effectiveness of the new materials and to identify areas where revision is needed.

When the new curriculum has been fully developed, evaluation makes it possible to determine the degree to which the new curriculum is effective in meeting the desired objectives for which it was designed.

Curriculum development is not simply a matter of devising new syllabuses, but rather there is a link between the study of the content of the curriculum and the methods of testing and presenting the content. Entwistle and Nisbet (1973) explain the importance of testing in any curriculum development:

> New materials and methods are tested in pilot schools and these development stages are (or should

> be) followed by a systematic evaluation to assess how effectively the objectives have been achieved by the content and methods which have been adopted. (p.264).

The Qatari situation lacks the indispensable duty to link curriculum development to newly adopted testing techniques to enable curriculum developers to determine the desired effectiveness of the applied curriculum. Many attempts have been made to develop textbooks, but less attempts have been suggested to improve testing.

3.2.3 Placement and Grouping:

Placement testing is concerned with students' entry performance and typically focuses on answers to a set of questions. These questions have been mentioned by Gronlund (1981).

1- Does the pupil possess the knowledge and skills needed to begin the planned instruction?
2- To what extent has the pupil already mastered the objectives of the planned instruction?
3- To what extent do the pupils' interest, work habits and personality characteristics indicate that one mode of instruction might be better than another? (P. 17).

The goal of placement testing is to decide the position in the instructional sequence and the kind of instruction that is needed by the students to benefit the most. Test results are

used to place individuals in the most appropriate classes. When students apply for a course, only tests can tell us where exactly to place them by measuring the degree of proficiency to assign them for specific activities on the basis of their current level of competence. In designing a test for placement, the test developer may choose to base the test content either on language proficiency or on the learning objectives of the syllabus to be taken. If the test is based on language proficiency, the test may be developed to determine placement according to a norming procedure. If, on the other hand, the objectives of the program are clearly defined, the test developer is more likely to develop a multi-level test based on the content objectives of the program.

At Qatar University, there are two kinds of study as far as the English language is concerned. There is the English Department for those who intend to specialize in English. Those students are tested at the beginning of their academic year. The test that they must take can be called an entrance test (or readiness) to decide whether or not they should enter the program and whether or not they are ready for the intended instruction. Students are tested and the proposed number of students is decided in advance. Students who score high on this test are selected for the study. The second kind of study at the University is called "The English Language Teaching Unit". In this Unit, English is taught as a University

requirement only. In a word, we can say that there is no placement testing in Qatar either at the University, or at the school level. I mean to say that students are not tested to be classified into suitable courses but as a routine procedure to be accepted as university students.

3.2.4 Predicting Future Performance:

Generally speaking, test scores create the possibility of being able to predict individual future performance based on his/her test scores. At the end of each school year, administrators, parents and students are interested in deciding what kind of course the students will take in the following year. They all use the test scores as a clue to predict how well the students will perform in the coming course. In Qatar, this important role of testing is not clear since the students' scores are not real representation of their abilities. For instance, students in Secondary One who get high marks on the Maths test decide to go to the scientific section. When they are tested in the scientific section, they fail to achieve the passing mark on the Maths test. This means that their previous scores were not real representation of their abilities.

3.2.5. **Describing Realistic Standards of Achievement:**

Students usually desire to know how much they have achieved and where they stand among their classmates. Tests have the ability not only to provide the level of achievement, but also to establish exactly what a learner has or has not mastered.

Test scores may urge a student to compete with others and sometimes with himself. Low scores may motivate a student to double his effort to improve his own achievement, while high scores give him a feeling of satisfaction and confidence, his success will lead to further success.

3.2.6. **As a Diagnostic Procedure:**

Testing experts consider tests as a process through which teachers can evaluate their teaching, whether it was effective or not. After tests, teachers often ask themselves a lot of questions whose answers can make the final decision about improving the teaching process. These questions that teachers often ask are introduced by Madsen (1983):

> Have I been effective in my teaching?
> Are my lessons on the right level?
> Am I teaching some skills effectively, but others less effectively?
> Which points need reviewing? (P.5).

All those who are concerned with testing agree that testing may be used to enable teachers to identify specific areas of strengths and weaknesses so that they may plan the most appropriate remedial procedure to overcome the weaknesses and enforce the strengths. Through tests that teachers make, they can diagnose which aspects of the materials present particular difficulty for pupils. The classroom teacher needs to know if his students have mastered what he had taught them in order to use the most appropriate techniques accordingly. The classroom teacher diagnoses the weaknesses in the tested area and then make the suitable correction for these weaknesses and reinforces the strengths.

3.2.7. To Grade Students For Promotion:

One of the main reasons for testing is to grade students so that teachers can move them to the next higher class or retain them at their present level. Without testing, promotion will be impressionistic.

We have seen that testing plays an important part in the teaching-learning process. It provides a background for decision-making about students and their attainment in a certain subject during a certain period of time. Testing is an indicator of the degree we are progressing towards course objectives. Feedback on the effectiveness of students' learning

is generally of interest to both teachers and students for making decisions regarding appropriate modifications in the instructional procedures and learning activities. Finally, testing is the basis for the promotion of students to the next stage of learning.

Theoretically speaking, the Qatari situation as far as the purposes of testing are concerned has many purposes of testing, but only one of these is achieved on the real schools levels. It is clearly identified that testing in the Qatari schools has just one purpose which is to promote students to a higher position of learning. Teachers test their students just to assign marks to a given situation which is called testing. Teachers ignore the fact that testing has many different purposes if it is done properly and adequately. They are also unaware of the fact that testing has something to do with our ways of presentation and it is firmly connected with what the students have learned. The present situation has nothing to do with refinement and aspiration to a better teaching and testing. What happens in testing vocabulary, for example, does not fulfil the possible purpose. Teachers have four or six marks that are assigned for testing vocabulary. They just give four or six isolated words and ask their students to give the equivalent in their native language and vice versa. In fact, testing vocabulary is meant to give the impression that the students understand the target word not in isolation, but to

have the ability to carry out orders that require either a simple physical action like pointing at something or a very simple answer such as "yes" or "no". In good teaching, teachers should not recommend having students simply memorize lists of words. On the contrary, they should teach the students to find the meaning of words through the context of the sentence. Since words are not taught in isolation, words should not be tested in isolation.

In a word, purposes of testing should be studied and understood in a way that can make a change for the better. The testing situation in the Qatari schools lacks the ability to set realistic standards of achievement for groups or individuals. Present testing procedures are not intended to diagnose weaknesses and re-enforce strengths. They are just routine procedures that teachers are asked to do now and then.

3.3.1 Different Types of Test:

We have seen in the previous discussion that testing has many different purposes. To achieve these purposes, we cannot use only one type of test to get what we aim at.

To classify different types of language test, test developers should consider many features that help in this classification. Language test developers and users are

frequently faced with questions regarding what type of test would be most appropriate for a given situation. Bachman (1990) comments saying:

> In discussions of language testing one often hears questions such as, "Should we use a norm-referenced or an achievement test?" or "Should we use both a diagnostic and a proficiency test in our program?" Such uses of labels for describing test types often raise more questions than they answer. How are norm-referenced tests different from achievement tests? Cannot proficiency tests be used for diagnosis? Questions like these imply comparisons that are like the proverbial question, "Which are better, apples or oranges ?" (P. 70)

The previous quotation implies the idea that types of test may be classified according to different features. These features are given by Bachman (1990):

> Language tests can be classified according to five distinctive features: the purpose, or use, for which they are intended; the content upon which they are based; the frame of reference within which their results are to be interpreted; the way in which they are scored, and the specific technique or method they employ. (P. 70)

Comprehensive testing always requires a variety of testing types. No single technique has been proved to be completely adequate and reliable to appraise students' progress to achieve the well defined and important outcomes of systematic instruction. An achievement test, for example, is commonly used to provide the tester with a true picture of a variety of learning objectives such as concepts, knowledge of terms. etc. While a diagnostic test is constructed and administered to

diagnose certain weaknesses and strengths. When results have been obtained, remedial instructions must be applied to reinforce strengths and get rid of weaknesses.

Karmel (1978) says:

> A school should administer many different types of tests because no one test can measure all the varied facets of a child's ability, interests and personality. (P. 4)

3.3.2 Achievement Testing:

The most important aim of achievement testing is to identify and define the learning outcomes that are to be measured. They measure specific types of performance that students are expected to demonstrate at the end of instruction. Gronlund (1982) also gives the following definition of the term.

> An achievement test is a systematic procedure for determining the amount a student has learned. (P. 1)

Brown (1981) says:

> Achievement test is a test designed to measure the extent to which a person has acquired certain information or mastered certain skills, usually (though not necessary) as the result of specific instruction. (P. 208)

This type of testing arose from practical school situations. There have been many studies about this type of testing, which led to the objective trend in testing, but the beginning of objective testing is difficult to trace. Adams (1964) says:

> When one considers the widespread use of standardized tests today, it is difficult to realize the youth of objective testing. (P. 428)

The most important development of the early 1920s was the birth of organizing tests into batteries. In 1922 the Stanford Achievement Test appeared. Since then, achievement tests has been drawn into playing a basic role in classroom activities.

When dealing with achievement testing, two kinds can be discussed. The first is the final achievement test which can be administered at the end of a course of study. This test should be based on a detailed course syllabus or the books used in the course of study. In this respect, the test only contains what it is thought that the students have actually encountered. The purpose of this is to check if students have mastered what has been taught to them in order to move them to the next objective. Testing at the end of instruction also helps teachers to adjust their previous strategies of instruction and students can know the level of their abilities for future

performance. Gronlund (1982) summarizes the goals of achievement testing saying:

> The results of achievement tests can be used to evaluate various aspects of the instructional process. They can help determine the extent to which the instructional objectives were realistic. Whether the methods and materials of instruction were appropriate, and how well the learning experiences were sequenced. Test results not only reveal the learning weaknesses of individual students, but when the results of the group are considered as a whole, they can also reveal weaknesses of instruction. (P. 7)

Achievement tests can be designed to test a variety of learning outcomes, such as the knowledge of specific facts, the knowledge of terms and understanding of concepts and principles.

The second is the progress achievement test which is intended to measure the progress that students are making towards the course objectives. This test can be based on short-term objectives. The progress test can be looked at as a unit-test which gives the indication that what has been taught, has been mastered by the students.

Classroom teachers are always involved in the preparation of achievement tests. These tests measure a representative sample of learning objectives. Achievement tests help in having an idea of what has been instructed and the

degree of students' mastery of these materials. Gronlund (1982) summarizes the basic principles of achievement tests:

1. Achievement tests should measure clearly defined learning outcomes that are with harmony with the instructional objectives.
2. Achievement tests should measure a representative sample of the learning tasks included in the instruction.
3. Achievement tests should include the types of items that are most appropriate for measuring the desired learning outcomes.
4. Achievement tests should fit the particular uses that will be made of the results.
5. Achievement tests should be as reliable as possible and should then be interpreted with caution.
6. Achievement tests should improve students' learning. (PP. 8-13)

The aim of achievement tests is to help both the classroom teacher and the students. The classroom teacher feels, after the test, job satisfaction and the students are encouraged to make the necessary progress. Information that can be derived from the test will inform the test user about those who can pass the test and others who cannot. In this respect, those who pass can be moved to a higher stage of learning (see figure 3.1)).

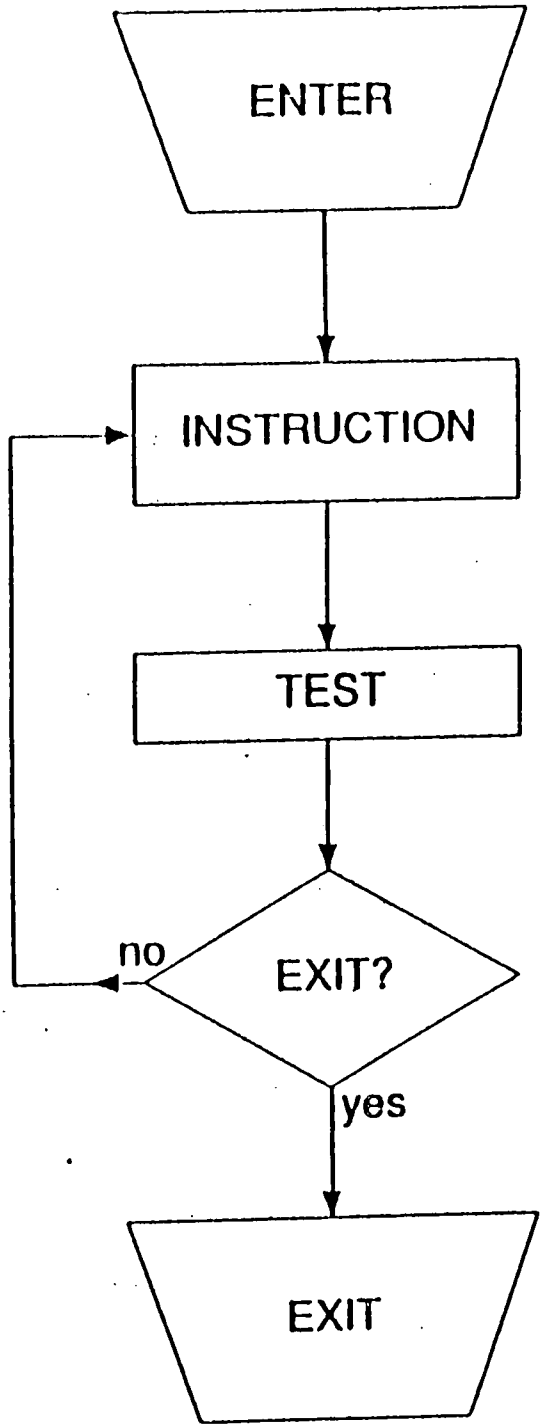


Figure 3. 1 *(Upshur 1970)*

and those who fail the test should be instructed again and prepared to take the test for the second time (see figure 3.2)

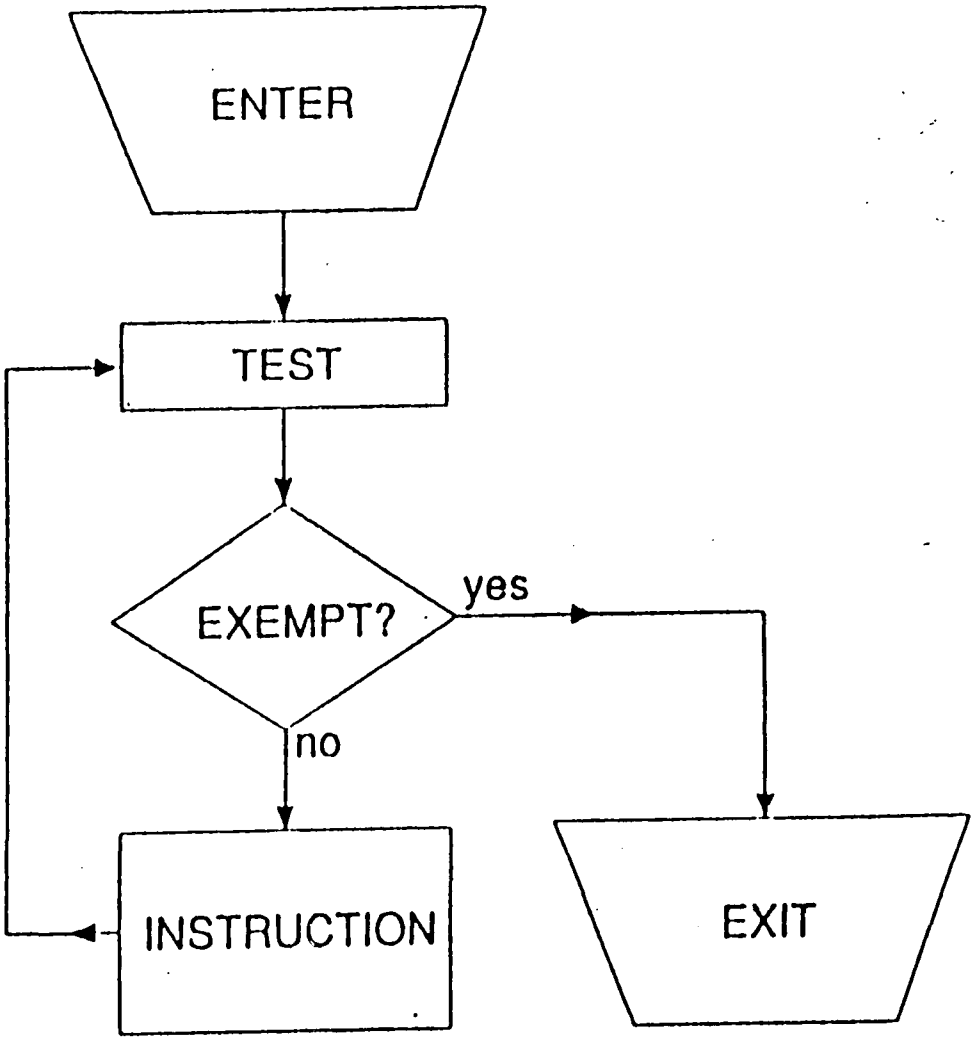


Figure 3.2 *(Upshur 1973)*

3.3.3 Standardized Tests:

Standardized tests have an important function in testing and education. They include many types, but they all share the following characteristics that were given by Brown (1981):

1- The tests are carefully constructed with extensive tryout, analysis and revision of items before the final form of the test is prepared.
2- Explicit procedures for administering and scoring the test are developed.
3- Normative data, derived from the administration of the test to carefully selected group, are available as aids in the interpretation of scores. (P. 141)

Bagnato (1981) defines standardized tests as:

> A systematic sample of performance obtained under prescribed conditions, scored according to definite rules and capable of evaluation by reference to normative information (P 638)

Hopkins and Antes (1985) add new features to the term when they define it:

> A commercially printed test for which content has been selected and checked empirically. The test is standardized so that the administration and scoring procedures are the same for all test takers. (P 479)

Thus, standardized tests are ready made as a result of research and development in a statistically accepted manner, their construction is produced by experienced personnel to a highly technical standard. This type of testing is primarily designed to measure overall achievement. It seems to me that standardized tests cannot be used in the classroom to test achievement of immediate and specific instructional objectives because teachers do not practise this kind of testing. These tests include a fixed set of test items designed to measure a clearly defined achievement domain. They have specific directions for administration and scoring the test

3.3.4 Teacher-Made Tests:

Measuring pupils' progress is a major aspect of the teacher's responsibility in the learning process because teachers have achieved a more significant role in education. These tests are most likely to be concentrated on, by teachers, because they have close relationship to classroom teaching. They can provide the classroom teacher with a true idea about the students' progress, his effective ways of teaching and about the degree of achievement of the defined objectives.

On the other hand, teacher-made tests constitute the basis for grading students, reporting to parents and provide the students with a feedback concerning the effectiveness of their

UNIV DUNELM

efforts to learn. Adams (1964) quotes Brownell who has developed several criteria for judging the worth of classroom tests in relation to the instructional process:

> Does the test elicit from the pupils the desired types of mental process?
> Does the test encourage the development of desirable study habits?
> Does the test lead to improved instructional practice? (P. 324)

Although achievement can be measured either by standardized tests or teacher-made tests, it is always preferable to be done by teachers because they are the only individuals who live the classroom's life moment-by-moment and they are fully involved in classroom activities. This means that teachers are in the best position to evaluate the situation. Teacher-made tests can provide true picture of the students' ability. Ebel (1965) explains the nature of classroom grades.

> Tests can, and often do, help teachers and professors to give more valid, reliable grades, because they are intended to summarize concisely a comprehensive evaluation of the student's achievement, because they are reported to the student and his parents to indicate the effectiveness of his efforts. (P. 6)

3.3.4.1 Advantages of Teacher-Made Tests:

Teacher-made tests, as any type of testing, have certain advantages that can be observed when putting the test into practice. Such advantages are:

1- Teacher-made tests are based on a representative sampling of the materials studied.
2- Teacher-made tests fit the content and objectives of a particular course better than would a test prepared by any one else.
3- When teaching and testing are in the hands of the same person, they are likely to be more effective because the teacher is closely acquainted with students' individuality, he can decide how to motivate them and when exactly to do it. Satterly (1985) gives the following strengths of teacher-made tests:

    1. They provide a closer fit between test and course content.
    2. Tests which correspond with units of a course provide a check of the effectiveness of teaching and learning.
    3. Teacher-made tests can provide a more extensive sampling of specific areas of learning.

4. Teacher-made tests provide information regarding a pupil's attainment relative to specific objectives which then provide a basis for report to parents concerning his/her educational progress. (P. 74)

3.3.4.2 Disadvantages of Teacher-Made Tests:

Teachers often reveal shortcomings in their construction, use, development and scoring of tests because most of them are not skilled in the art of testing. Since I am a teacher of English as a foreign language, I can list the following disadvantages:

1- Teacher-made tests are usually short and poorly constructed.
2- Teacher-made tests do not make distinction between basic and trivial objectives. This means that teachers cannot sometimes distinguish between the important areas to be tested from the unimportant ones.
3- Teachers do not know the effective role that can be played by their tests in directing learning in the classroom, so they test just for the sake of testing.
4- Most teachers copy bookish terms in the test instructions. They never try to invent new ones, that is to say, they copy what most textbooks give at the

end of each unit as a sample test.

5- Teachers sometimes use items that may be of great help to answer another item. The information in question "A" may be used to answer question "B".

6- Teachers are sometimes described as conservative, that is, they are happier with familiar types of testing than with new ones.

7- Teacher-made tests are not reliable.

Ebel (1965) explains the reason:

> Many teachers have never checked on the reliability of any of their tests and may not even have planned those tests purposely to make them as reliable as possible. (P.14)

Finally, as a teacher of English as a foreign language, I can say that teachers of English in some countries are to be freed from blame in the sense that they lack knowledge about teaching and testing for many reasons. When teachers were students at the university level, they have not been instructed how to teach and test the language they learn. I remember when I began to teach in Qatar, The Ministry of Education has introduced us to a course of orientation about how to teach and test the communicative approach. The time that was spent on this course lasted only about ten days which was not adequate to introduce the teachers to a complete and new course. So teachers always depend on personal initiative to obtain the

needed knowledge either in teaching or in testing. Thus, I think that if teachers do things and succeed, they should be praised

In a word, teachers are responsible for evaluating their teaching and sometimes reporting the progress that has been made. In Qatar, teachers do not feel this responsibility. On the contrary, they still think that testing is just a routine procedure to be done monthly. They never try to evaluate their tests at all. So, students are still traditionally tested.

3.3.5 Norm-Referenced Testing:

Both types, criterion-referenced and norm- referenced tests, are of great importance since each provides unique information concerning student achievement.

In this section, our discussion will be restricted to norm-referenced tests, and a full comparison of the two types will be given at the end of this chapter.

The term, norm-referenced tests is composed of two parts, the first which is 'norm' needs clarification, then we shall proceed to define the term as a complete form. The concise Dictionary of Education (1982) defines norms as:

> Extensive statistical findings that document how representative groups of students have actually performed on a standardized test, for adequate score information, norms often need to be provided on bases

> that are nation-wide, state-wide, school-system-wide and comparable to the students concerned in other essential respected such as sociobiological status. (P. 154)

Adams (1964) clarifies the meaning of norms by saying:

> Norms are statistics that describe the test performance of specified groups, such as pupils of various ages or grades in the standardization group for a test. Norms are often assumed to be representative of some large population of pupils in the country as a whole. Norms are descriptive of average. (P. 643)

Tuckman (1975) explains the function of norms in testing:

> Norms are statistics that supply a frame of reference by which meaning may be given to obtained test scores. Norms are based upon the actual performance of pupils of various grades or ages in the standardization group for a test. Since norms represent average or typical performance, they should not be regarded as standards or as universally desirable levels of attainment. (P. 481)

It is obvious now that norms play an important role in interpreting test scores. All tests users know that raw scores on a test have no meaning unless interpreted in relation to some reference. Norms provide a reference against which to compare performances and indicate where students in the same group stand in relation to each other. Norm- referenced tests tell test users how an individual compares with other

persons who have taken the same test. Gronlund (1982) restricts the use of norm-referenced tests:

> Since norm-referenced tests are designed to indicate how an individual's test performance compares to that of others. (P. 113)

Norm-referenced tests are of great importance to classroom teachers when they wish to compare their students' performances. Norm-referenced tests are used to fulfil certain purposes, some of which were given by Roid and Haladyna (1982)

> The purpose of a norm based comparison is presumably to: (a) group students for instruction. (2) assign honours for high achievement. (c) select for special programs. (d) evaluate an instructional program. (P. 27)

Classroom teachers sometimes ask which type they should use. The correct answer depends on the purpose of the test. If teachers are planning to make decisions about individuals' status with reference to some criterion they can use criterion-referenced tests. If they need to make decisions about individuals' status relative to performance of others, they can use norm-referenced tests. In fact, the two types are needed to make decisions about individuals. The purpose of the test decides the type to be used.

3.3.5.1 Advantages of Norm-Referenced Tests:

Most of the advantages of norm-referenced tests are directly tied to their ability to tell us about the student position among others. Hopkins (1985) lists the strengths of norm-referenced tests as follows:

> 1- NRM can be used to measure attributes that relate to the cognitive domain where each student is allowed and encouraged to learn at maximum potential, for this reason norm-referenced tests are needed for nearly all classrooms.
> 2- NRM is appropriate for interpreting performance at the higher levels of the affective and psychomotor domains. As performance becomes complex, comparison to others becomes increasingly important in interpretation of individuals performance. (P. 333)

This type of testing has the ability to discriminate between individuals and points out the weaknesses of both students' achievement and teachers' presentation. This type can also help when we intend to select students. The employment of both types provides more flexibility and understanding of student progress and the level of instruction.

3.3.5.2 Disadvantages of Norm-Referenced Tests:

It is clear that norm-referenced tests are unable to provide information about the level of achievement in direct measures. Hopkins (1985) says:

> 1- Norm - referenced measurement is not appropriate for measuring mastery of certain materials and skills.
> 2- Norm - referenced measurement does not indicate specific tasks a student can perform and does not allow direct interpretation of performance. (PP.333-334)

3.3.6. Criterion-Referenced Testing:

Interpretation of tests scores is a complicated process and at the same time a necessary procedure to teachers to give meaning to a raw score either by converting it into a description of the specific task that the pupil can perform (criterion-referenced) or converting it into some type of derived score that indicates the pupil's relative position in a clearly defined group (norm-referenced). Both criterion-referenced and norm-referenced are two ways of interpreting test scores. I shall draw a comparison between these two types. The strengths and weaknesses of each type will also be highlighted. Purposes of each type will be within the scope of our coming discussion.

Criterion-referenced testing is designed to measure defined goals of learning. Its aim is to provide a useful description of students' attainment that is relative to stated outcomes.

The first use of the term was either in 1962 or 1963. Riod and Haladyna (1982) say:

> Glaser (1963) was the first to use the term criterion-referenced. (P. 7)

Brown (1981) does not agree with Riod and Haladyna that 1963 was the beginning of criterion-referenced testing:

> 1962 was the first use of criterion-referenced measures. (P. 23)

Since Glaser's use of criterion-referenced tests, the literature has accumulated over 600 references to criterion-referenced testing.

Testing individual performance is a topic of common concern among those who are interested in business, industrial affairs and education. This type of testing was developed to answer teachers' questions about individual performance. Teachers and educators looked at norm-referenced tests as a limited type of testing. Satterly (1985) describes the birth of this type of testing.

> Criterion-referenced tests have been developed in response to some of the limitations of norm-referenced testing. (P. 48)

Satterly (1985) goes on to say:

> Norm-referenced tests are obviously unable to provide much of the necessary information for a teacher who wants to give children as much time as is practicable to master a given objective. (P. 49)

Karmel (1978) expresses educators' views about the limitation of norm-referenced tests.

> Many educators have come to the conclusion that comparison with others is not always the best way to evaluate an individual. Therefore, for several years the trend has been to develop measurement techniques in which individuals are evaluated in relation to self achievement or to an unchanged absolute standard. (P. 377)

What interests teachers is not the students' position among others, rather they are more interested in students' attainment according to a previously established standards.

The appearance of criterion-referenced testing marked a new interest in providing information about individuals. During the seventies, interest in criterion testing has been increased and many books and articles have been written about the term since it was popularized in 1963. It is really a recent

addition to the area of individual performance which allows every pupil to reach an acceptable level of performance without any competitive comparisons which are considered to be of a detrimental effects. Karmel (1978) says:

> Experts speculate that comparisons are particularly damaging for the young child. (P.377)

Since the appearance of criterion-referenced tests, a number of definitions have been offered. In this chapter, different definitions have been quoted to highlight the term and its use in education:

1- The International Encyclopedia of Education (1985) says:

> Criterion-referenced tests are constructed to permit the interpretation of examinee test performance in relation to a set of well-defined competences. (P. 1108)

2- Brown (1981) gives a general definition of the term:

> Assessment that provides information about specific knowledge and abilities of pupils through their performance on various tasks that are interpretable in terms of what the pupils know or can do, without reference to the performance of others. (P. 4)

3- Billing (1973):

> Test designed to measure the degree to which learning programme attains its objectives. (P. 156)

4- Bugnato and Neisworth (1981)

> Instrument designed to identify the presence or absence of specific skills in terms of absolute levels of mastery. (P. 324)

It is clear that all the given definitions concentrate on skills and abilities that are related to the individual's attainment.

3.3.6.1 When to Use Criterion-Referenced Tests:

Criterion-referenced tests are used and most applicable in the area of skill testing. Usually, criterion-referenced testing is the most prevalent kind which is used in schools. Karmel (1978) says:

> Each time a teacher gives a quiz or examination, she or he is probably using criterion-referenced testing, or measuring actual performance. (P. 379)

This type of testing can also be used to measure behaviours such as problem solving, there is also an interest in making greater use of criterion-referenced testing in school attainment whether it is internal or external testing. Tuckman (1975) outlines the use of criterion-referenced testing.

> Criterion-referenced is used by (1) generating or selecting a set of objectives representing the desirable performance outcomes of instruction, (2) designing or finding items to adequately measure each objective, (3) presenting acceptable performance levels. (P. 296)

I have found that different testers give different views about criterion-referenced testing. Some testers are enthusiastic when they talk about criterion-referenced testing in that they approve their use of criterion-referenced testing in the most difficult area: the cognitive domain. Hopkins (1985) says:

> CRM can be used to measure attributes of the cognitive domain where mastery of certain materials and skills is expected. For this reason, some criterion-referenced tests are needed for nearly all classrooms. (P. 327)

Others are less enthusiastic when they talk about the use of the term. Brown (1981) has quoted Henrysson (1974) to tell us that criterion-referenced tests are only used to measure low levels of activities.

> Criterion-referenced tests are only suitable for subject areas with a simple structure. (P. 10)

While Tuckman (1975) has quoted Ebel (1970) to say:

> Criterion-referencing is most applicable to the measurement of complex behaviour such as thinking and problem solving. (P. 296)

Criterion-referenced tests are mostly used in school systems to measure students' individual achievement. Teachers who are responsible for appraising classroom activities are sometimes faced with individuals who need special treatment on the basis of their unsatisfactory attainment. In this situation, teachers can use criterion-referenced tests to diagnose students weaknesses and decide a remedial individualized instruction.

Criterion-referenced tests can also help in making instructional decisions since systematic instruction requires close and continuous monitoring of students progress. This type of testing can also be used in the traditional meaning of testing, that is; to increase student motivation, assign grades, evaluate instructions or select students for special programs. Roid and Haladyna (1982) point out the new uses of tests in systematic instruction:

> First and foremost, tests are used to certify that student achievement has been satisfactory. (P. 26)

The International Encyclopedia of Education (1985) mentions the following uses of criterion-referenced tests:

> In relation to the competences there are three common uses for criterion-referenced tests scores: (a) to describe examinee performance, (b) to assign examinees to mastery states, (c) to describe performance of specific groups. (P. 1108)

It is obvious now that criterion-referenced tests are commonly used in school systems. Classroom teachers use these tests to be able to locate their students correctly in school programs, to diagnose students' learning deficiencies, and to grade students on the educational ladder.

One of the diagnostic features that are produced after the test is that students can evaluate their performance with reference to specific objectives. Test scores provide a basis for making decisions about learning and teaching. A good criterion-referenced test tells us about the level of our students' achievement and also about our way of presentation. Tuckman (1975) explains this diagnostic value of criterion-referenced tests.

> Where instruction is of a group nature, test results must be applicable to judgments of group progress. If group success has largely been attained, then

> instruction on new materials can begin. If group gains have been minimal, then remedial instruction should be provided before instruction can progress to new areas. (P. 300)

In short, criterion-referenced tests can provide us with the necessary information that help us in making a final decision about classroom situation. They also tell us whether adjustment or remedial efforts are needed or not.

3.3.6.2. Criterion-Referenced Test Construction:

It is common in test construction to define the objectives of the course as a first step. As for criterion-referenced tests, it is essential to specify clearly the behaviours defining each competence that is to be tested.

In the case of teachers of English as a foreign language, teachers should outline the course content in terms of vocabulary, grammatical structure such as Past Continuous and the Simple Past, functions such as comparing, suggesting, asking polite questions, and the aims to be taught in each step in the teaching process. Such outlines help the classroom teachers to know when constructing items for a new test, whether or not certain expressions have been already tested, or have been overused, while other items have been completely left

out. Tuckman (1975) suggests steps that we should go through when constructing criterion-referenced tests. The steps are:

> (1) Prepare a content outline listing the skills and knowledge that the test is an attempt to measure, (2) Identify the performance, (3) Identify the domain that each objective defines. (4) Validate the fact that the skills and knowledge measured by the test are in fact prerequisite to the performance objectives. (P. 295)

To clarify the idea of making an outline of course objectives to be tested, I shall give an example. Before we construct the test the following outline of objectives will certainly help.

For example, in Qatar, English is taught from primary 5. Let us call this level, the lower level and consider the tasks which are set for this level. The following tasks are quoted from English Language Syllabus, Curriculum and Textbook Department, Ministry of Education, (1988)

3.3.6.3 Tasks for Level 1. (lower).

3.3.6.3.1 Listening:

By the end of the Primary stage the pupils should be able to:

1. Listen to English recorded or on educational and recreational tapes and develop the desire to do so.
2. Follow a short dialogue and demonstrate his understanding of it by responding orally, in writing or by action. The topics have to be restricted to the pupil's environment. e.g. classroom, school, family-life, etc.
3. Respond correctly and appropriately to the classroom instructions given by the teacher in simple language.
4. Listen to a short conversation on a familiar topic and be able to identify the main ideas, the situation and the role of the participants, the recording to be clear and slower than normal speech but without distortion of natural pronunciation.
5. Respond to the language (vocabulary and functions) related to this stage, as specified in Components of the English Syllabus by the Gulf Arab States Educational Research Centre (GASERC).
6. Understand the grammar rules needed to comprehend the audio-lingual materials presented to the pupils.

3.3.6.3.2 Speaking:

By the end of the Primary stage the pupil should be able to:

1. Take part in the learning activities related to the listening skill and demonstrate his understanding by responding orally.
2. Pronounce all the language (words and structures) in the syllabus clearly using correct and appropriate intonation.
3. Ask questions and request help from the teacher or from classmates using the appropriate forms.
4. Respond in English to a limited numbers of question types by using the correct short answer forms.
5. Perform some short role-plays and rhymes using the tape of the teacher as a model.
6. Use the language (functions and vocabulary) related to this stage as specified in Components of the English Syllabus by Gulf Arab States Educational Research Centre (GASERC).

3.3.6.3.3 Reading:

By the end of the Primary stage the pupil should be able to:

1. Read the alphabet in capital and small letters, to distinguish between the two, to recite the alphabet in order and put any group of letters into alphabetic

order.

2. Read the numbers from one to 1000 and distinguish between them.
3. Read the contents of the textbook assigned for the Primary stage: all words, sentences and discourses as well as flashcards and transparencies.
4. Read shop signs, traffic signs, warning notices etc.,commonly seen in Qatar.
5. Read the books and magazines assigned for this level.
6. Read unfamiliar words by using English spelling and pronunciation rules.
7. Look up the meaning of words in the glossary at the end of the pupil's Book.
8. Read all the language (words and structures) related to this stage as specified in Components of the English Syllabus.
9. Understand the grammar rules needed to comprehend the assigned texts.

3.3.6.3.4 Writing:

By the end of the Primary stage the pupil should be able to:

1. Write all the letters of the alphabet and the numbers clearly, tidily, and correctly.

2. Copy sentences and discourses accurately.
3. Answer, in writing, a spoken or written question.
4. Write one or two sentences describing familiar things with the help of a model.
5. Write with correct spelling, most of the words in the textbook.
6. Write most of the sentences types for this stage as specified in Components of the English Syllabus.
7. Use correct punctuation: capital and small letters, commas, full stops, question and exclamation marks and apostrophes

These tasks can provide guidance about the language skills to be included in language tests. The limited nature of the educational outcomes and specification of the learning tasks clarify the use of testing.

Task analysis is a necessary prerequisite because such an analysis enables identification of the critical elements required for successful task performance. When the objectives have been carefully defined, we move to the next step which is to select the most appropriate technique to test these objectives. We should always keep in mind that testing is a process which becomes most effective when it is used on the basis of sound operational principles.

When the objectives or aspects of pupil performance have

been defined, the testing technique should be selected. Gronlund (1981) tells us the steps to be followed in constructing a test.

> 1- Determining and clarifying what is to be evaluated always has priority in the examination process.
> 2- Evaluation techniques should be selected in terms of the purposes to be served.
> 3- Comprehensive evaluation requires a variety of evaluation techniques. (PP. 22-23)

The technique that should be selected in criterion-referenced test has special characteristics that were given by Hopkins (1985):

> The thrust of test construction for criterion-referencing is toward building a device that will generate information about whether a student can or cannot do those things expected from all students. (P. 32)

Criterion-referenced tests are constructed so that the results can be interpreted directly in terms of the specific tasks and skills that the pupils can demonstrate. If we want the test to be valid and provide valid description, we should select each objective and treat it separately. Although this technique needs a lot of time from teachers to construct and administer, some testers prefer to use it. Gronlund (1981) has quoted Popham who stresses this kind of separation.

> Separate test specification are needed for each set of items. (P. 71)

Klausmeier and Goodwin (1971) agree with Gronlund that objectives should be tested separately:

> Such a test is usually short, as few as 10 items, because it measures a single objective or small number of closely related objectives. (P. 450)

Swezey (1981) also agrees with the previously quoted quotations:

> Compound objectives also must be broken into unitary objectives and even further, into their components, before proceeding with test item development. (P. 32)

When the previously discussed steps are taken, the teacher is in a position to start constructing the target test. Users of criterion-referenced tests should not stick to just one procedure when they intend to construct their tests to diagnose strengths and weaknesses of teaching-learning process. Swezey (1981) argues:

> There is no single correct way to construct a criterion-referenced performance test. (P. 15)

Criterion-referenced tests may be attractive to classroom teachers because these tests are easy to construct and suitable to learning tasks whether long or short in the learning process.

3.3.6.4. Characteristics of Criterion-Referenced Testing:

The most important characteristics of a test are reliability and validity. Now, the discussion will proceed to deal with these two characteristics with relation to criterion -referenced testing.

Reliability refers to the consistency of test scores over different occasions. On the other hand we should not expect all test scores to be perfectly consistent since these scores may be affected by many factors. To determine reliability we need to obtain two sets of scores and then compare them, but this is not always possible because test conditions are not always the same.

Criterion-referenced test reliability is still new. Until recently a number of authors did not know how to establish criterion-referenced test reliability. Swezey (1981) explains the difficulty of reliability establishment:

> Most techniques for assessing criterion-referenced reliability are exploratory, and are either not fully

> developed or are based on questionable assumptions (P. 143)

Satterly (1985) agrees with Swezey that it is still difficult to establish criterion-referenced reliability:

> There is as yet no universally agreed method for estimating the reliability of criterion-referenced tests. (P. 218)

Gronlund (1981) stresses the novelty of criterion-referenced reliability:

> A number of statistical procedures have been proposed for estimating the reliability of criterion-referenced tests, but methods are relatively new and still in the developmental stage. (P. 111)

If testing theories are not able to provide classroom teachers with a satisfactory method for establishing the reliability of criterion-referenced test what can teachers do to estimate their tests' reliability?. We should always keep in mind that where decisions are to be made about individuals, then reliability should be established. Gronlund (1981) suggests a simple method of estimating criterion-referenced reliability:

> For most classroom use of criterion-referenced mastery tests some relatively simple method of

> estimating reliability may be sufficient. Because in mastery testing we are primarily interested in classifying pupils as masters or nonmasters, the consistency of our mastery-nonmastery decisions should be our main concern. This type of reliability can be readily determined by computing the percentage of consistent decisions over two equivalent forms of the test. (P. 112)

Test reliability is of great importance for teachers' assessment of pupils. If we expect that tests provide us with information which will be used to evaluate pupils and make decisions about them, then we are in need of knowing how much confidence can be placed in the test results.

The simplest advice that can, at present, be offered is to interpret the concept of reliability of criterion-referenced tests in terms of the consistency with which a test makes a decision about the mastery state of the pupils taking the test.

We say that a criterion-referenced test is reliable if examinees who pass it on one administration, pass it on a repeated administration, provided no instruction is given between the two administrations. Validity of criterion-referenced tests has not received as much attention as reliability, although it is commonly agreed that it is necessary to this type of testing. Brown (1981) quotes Hambelton who said:

> The usefulness of any of the applications of criterion-referenced tests will depend on the validity of the intended interpretations of the criterion referenced test scores. (P. 54)

Validity is concerned with establishing that a test measures what it claims to measure. Validity has many types, the most important for criterion-referenced test is content validity which means according to Gronlund (1981)

> How adequately the sample of items represents the domain of achievement. (P. 85)

Brown (1981) stresses the importance of content validity in criterion-referencing.

> Most writing on validity in relation to criterion-referencing has concentrated on content validity. (P. 55).

She has also quoted Popham and Husek (1969) who both agree with Brown on the relevance and importance of content validity.

> Criterion-reference measures are validated primarily in terms of the adequacy with which they represent the criteria. Therefore, content validity approaches are more suited to such tests. (P. 55).

Swezey (1981) also gives priority to content validity.

> It is generally agreed that content validity is of paramount concern in criterion-referenced measurement. (P. 149).

Since teachers are concerned with classroom achievement, they should know that the main reason for classroom testing is to make sure to what extent the pupils have mastered what they have been taught. When teachers construct a criterion-referenced test, they need to know how well the items included in the test match the objectives of the curriculum and the real content of the teaching experiences. Knowing the importance and relevance of validity and its use in classroom measurement are not enough. Teachers should also know how to judge content validity. Swezey (1981) explains how we can judge this type of validity.

> A criterion-referenced test may be presumed content valid if all test items are carefully derived from the required performances, conditions, and standards specified in the objectives and if the sample of test items appropriately represent the objectives. (P. 149)

3.3.6.4.1 Advantages of Criterion-Referenced Tests:

When we construct criterion-referenced tests, we expect that interpretation of test scores will indicate precisely the level of each student knowledge in terms of specific objectives since these tests are originally planned to measure what has

been taught in the classroom.

Criterion-referenced tests have the advantage of being helpful in making teachers able to reach individual plans about their students and to make sure that what has been taught in the classroom has been mastered by the students. Hopkins (1985) summarizes some of the advantages of criterion-referenced tests:

> CRM can be used to measure attributes of the cognitive domain where mastery of certain materials and skills is expected. For this reason, some criterion-referenced tests are needed for nearly all classrooms. Criterion-Referenced measurement allows direct interpretation of the measure of performance. The interpretation indicates what specific tasks a student can perform (P. 327)

Test interpretation can point out lacking skills and knowledge in students' performance. Such interpretation will certainly help teachers in constructing direct plans to face learning deficiencies. Bush and Waugh (1982) explain the advantageous aspects of criterion-referenced tests.

> Criterion-referenced tests are particularly useful as guidelines for assisting the teacher in making individual plans for children. Since they are geared more specifically toward the academic material taught in the classroom, the teacher can learn which skill or which set of knowledge is lacking in a child's achievement. (P. 31)

Criterion-referenced tests are not only used on individual levels, rather they are some times used to evaluate school programs. This diagnostic feature of criterion-referenced tests creates the possibility of remediation and change or at least adjustment to move safely. Hambelton (1985) agrees that criterion-referenced tests have this diagnostic feature.

> Criterion-referenced tests results are also being used to evaluate various school programs. (P. 1111)

She has also quoted KoseKoff and Fink (1976) who say that.

> Such tests are not appropriate for use in large-scale evaluations of the effectiveness of programmes. (P. 9)

3.3.6.4.2.Disadvantages of Criterion-Referenced Tests:

Most of the disadvantages of criterion-referenced tests are tied to one deficiency which can be put like this: criterion-referenced tests are unable to help teachers to compare a student with another. Hopkins (1985) lists the weaknesses of criterion-referenced tests as follows:

> (1) Scores from CRM do not indicate a student's level of achievement relative to peers.
> (2) Criterion-referenced tests cannot be used to measure large portions of subject matter.
> (3) Use of criterion-referenced tests exclusively

does not allow a teacher to compare a student with other student performance. (P. 328)

I think that the weaknesses I have quoted have no justification since criterion-referenced tests are not primarily constructed to indicate differences among students. If we want to reach such a position, another type can be used to determine the desired comparison among students performances, that is, norm-referenced testing.

In a word, criterion-referenced tests are of great importance to classroom teachers because their task is limited to the decision of the level of attainment and nothing to do with comparison. If comparison is needed, then, norm-referenced tests can be used.

3.3.7.1. Differences Between CR. and NR. Tests:

Each of the two types has an important role to play in providing information about students, but there are many differences between them. These differences are of great importance to classroom teachers and test users in general. The differences can be summarized as follows:

1- Norm-referenced test items are selected on the basis of how well they discriminate among students.

Criterion-referenced test items are selected on the basis of how well they represent specific learning tasks. Gronlund (1982) (P. 20)

2- Norm-referenced scores are compared to how the scores stand in relation to scores of others who have taken the test.

Criterion-referenced scores are compared to a predetermined criterion to judge whether the score should be considered as passing or failing. Hopkins (1985) (P. 342)

3- Norm-referenced tests are predicated on the assumption that the distribution of any trait measured will resemble a normal curve.

Criterion-referenced tests are not based on the normal distribution of a trait, but are concerned with the actual behaviours that a person can perform. Karmel(1978) (P.378).

4- Norm-referenced tests scores indicate the position of a child's score in a large normally distributed group of scores.

Criterion-referenced tests indicate the level of child's mastery of a given objective. It does not

refer to the distribution of scores by other pupils. Satterly (1985) (P. 54)

5-Norm-referenced tests are used to see who needs remedial work or special educations.
Criterion-referenced tests are used to see what kind of remedial action is necessary. Satterly (1985) (P. 55)

6- Norm-referenced reliability is estimated by using test-retest.
Criterion-referenced reliability is estimated by the consistency with which the set of items for an objective classifies pupils as masters or non- masters. Satterly (1985) (P. 55)

7- Norm-referenced validity is examined by reference to external criteria.
Criterion-referenced validity is examined chiefly by comparing the actual content of each item. Satterly (1985) (P. 55)

8- Norm-referenced tests' role in education is probably of greater interest to researchers than to teachers although teachers use them for selectio.
Criterion-referenced tests role in education is to

evaluate the absolute success of instruction. Satterly (1985) (P. 56)

Finally, the previously quoted differences of both norm-referenced and criterion-referenced lie in the respective purposes, types of scores, their interpretation, and the role they play in teaching.

3.3.8.1. Objective Testing:

In the following sections, objective and subjective testing will be discussed in detail since we do not want to limit ourselves to only one type because a good test is that which has the advantages of the two types. On the other hand, tests in Qatar did not treat the four skills on a balanced basis because the tests' constructors limited themselves to objective types and made no use of the other type. Byrd (1986) criticizes testing in Qatar saying:

> Examinations were usually restricted to reading and writing (ticking off or circling the correct answer, filling in blanks, answering questions), the observed tests did not treat oral or listening skills, failed to focus on reading skills, and never had students to write in any authentic genres. (P. 21)

Most of the tests that are in use in Qatar lack necessary emphasis on writing which is commonly considered the most

difficult skill in learning English. If teachers use only objective tests, they test recognition rather than language because the testees only recognize the answers and do not construct it. In subjective tests, the testees are asked to express their ideas as a result of a series of development that the learner should acquire during the learning process. Objective tests require the testees to choose their answers from a specified list of alternatives rather than creating them by themselves. The testees are to write down a word, a phrase, or sometimes a number. This type of test is widely used to measure, aptitude, intelligence and achievement. The objective test is so called because the scoring procedure is determined when the test items are written. The construction of such tests need patience, creative ability and knowledge of the group to be tested .

Objective tests, as other techniques, have advantages and disadvantages. Let us discuss them separately.

3.3.8.2. Advantages of Objective Testing:

1- Objective tests have a rapid ability to gather information.
2- Objective tests do not bring difficulty in scoring because the scoring procedure is determined when the test item is written.

3- Objective tests have no fixed procedure to write them, so teachers must spend time to construct them and this will help them to get a better understanding of the tested contents.
4- The skill needed in objective items construction can be acquired through experience and practice.

3.3.8.3. Disadvantages of Objective Testing:

1- Objective tests have been criticized because they place the testee in a position to recognize rather than construct the answer.
2- Objective tests place great emphasis on reading and less on writing.
3- Objective items encourage guessing.
4- They may include misleading questions.

3.3.8.4. Types of Objective Test Items:

Items can be classified either according to their form or function. Here, we are more concerned about the form. Objective test items have the following types:
True-false items, matching, multiple-choice and completion items.

3.3.8.4.1 True-False Items:

True-false items are presented to propose something and the student has to discuss it mentally. Care must be taken to select important propositions as a basis for the test items. Choppin (1985) gives a reliable description of true-false items and their use in testing.

> The true-false item represents a descriptive statement and requires the examinee to indicate whether he or she judges it to be true or false. (P.3617).

Satterly (1985) gives advice on writing true-false items: (freely adapted)

a- Avoid any source of ambiguity.
b- Avoid the use of negative statements.
c- Balance the number of true and false statements in the test but keep the length of statements roughly equal.
d- Avoid lifting statements verbatim from textbooks. (P. 90)

3.3.8.4.2.Advantages of True-False Items:

1- They are easy to read and understand.

2- They represent a simple way to test factual knowledge.

3- By using true-false items, the teacher can test a great deal of materials in a short time.

4- Hopkins (1985) gives the following advantage:

> Generally, students can respond to about two true-false items per minute. (P. 135)

5-Brown (1985) says:

> Although true-false items are looked on with disfavour by many teachers, they have many advantages. They are relatively easy to construct and can be scored rapidly and objectively. Generally the items are short, so a large number can be included on a test. (P. 52)

3.3.8.4.3.Disadvantages of True-False Items:

1- Among the serious limitations of true-false items is the limitation of the areas that can be evaluated by them.

2- Ebel (1965) says:

> True-false items have been criticized for triviality, for ambiguity, for encouragement of rote learning, for susceptibility to guessing and for exposing students to error instead of truth. (P. 146)

3- Satterly (1985) gives a general view of true-false items.

> The disadvantages of true-false items are probably more apparent than their advantages. The probability of obtaining a correct answer by guessing is high, the items are restricted in use. (P. 89)

4- Wood (1960) criticizes true-false items saying:

> One of its draw backs is that the item constructor is too likely simply to adopt verbatim statements from a textbook, with perhaps the inclusion of some negative terms to make some of the items false. Such a practice encourages rote memorization. (P.24)

5- Lindvall and Nitko (1975) speak about true-false scores:

> Scores from true-false tests are typically less reliable. (P. 55).

3.3.9.1. Matching Items:

Both matching items and multiple-choice items have a common set of alternatives, the matching test consists of two lists of phrases, the stem and the response, then choice is made by students, they must match the stem to the correct response. Wood (1960) warns teachers of faults in this type of construction:

> Two common faults in the construction of matching items warrant mention. One to have the same number of elements in each list. Since with such a set of items the typical procedure is to have each answer apply once only ... A second mistake that commonly appears in the use of matching items is to include far too many elements in both lists. (P. 27).

When a decision has been taken for writing matching items, some considerations should be taken into account. Brown (1981) provides us with these remarks:

> 1- All parts of the item should be homogeneous.
> 2- Limit the length of each list.
> 3- Each item should have only one correct answer. (P. 53)

3.3.9.2. Advantages of Matching Items:

1- Matching items give a great opportunity to the teacher to test associations of materials.

2- Guessing in reduced.

3- Short time is needed.

4- Easy to construct and score.

5- They are suitable to test definitions, names and places that cannot be measured by other procedure.

6- They are able to test scattered factual knowledge.

7- Satterly (1985) quotes Mehrens and Lehman (1978) who describe this type as:

> Most suitable for the 'Who', 'What', 'When', 'Where', types of learning and for any others which involve simple relationships. (P.99)

3.3.9.3. Disadvantages of Matching Items:

1- Matching items are very restricted in testing.

2- They fail to test instructional objectives.

In a word, the objective types of test are not as easy as we think. They impose certain measures on the item constructor. He should, first of all, know the limits of his subject, understands his students' abilities and specify clearly what to test and why. I think that all the objective items fail to encourage writing. They are constructed in a manner that does not require any writing effort, except just copying.

3.3.10.1. Multiple-Choice Items:

Multiple-choice items are currently the most highly regarded form of objective test item. The basic structure consists of two parts; the stem which represents the problem and a set of two or three options that represent the answer to the problem stated in the stem.

This type can be used to test many educational outcomes such as the ability to apply information, knowledge and reasonable judgement. Ebel (1965) summarizes the educational outcomes that can be tested by this type:

> They are adaptable to the measurement of most important educational outcomes, knowledge,

> understanding, and judgement ability to solve problems, to recommend appropriate action to make predictions. (P. 149)

When the teacher wants to write the items, certain suggestions may be useful to be taken into consideration, these suggestions were given by Lindvall and Nitko (1967)

> 1- Avoid the use of negative worded stems.
> 2- Make the responses as short as possible. (P. 58)

3.3.10.2. Advantages of multiple-Choice Items:

1- Multiple-choice items can be used to measure many educational outcomes.

2- Many items can be given on a test.

3- Brown (1981) says:

> Research has shown that multiple-choice items are more reliable and valid than other types of items. (P. 45)

4- Guessing is relatively reduced.

5- Objective scoring.

3.3.10.3. Disadvantages of Multiple-Choice Items:

In spite of the role they play in testing, multiple - choice items have not escaped the attention of critics. Some of the shortcomings are the following:

1- They are ambiguous.

2- They are not sharply discriminating.

3- There is a chance for guessing.

We conclude, that no type of testing can claim perfection.

3.3.11.1. Subjective Testing:

A subjective test requires the testees to express their ideas in a logical and competent way. On the other hand, requires the marker(s) to evaluate and not to follow a predetermined mark sheet. Subjective tests place a premium on the ability to produce rather than to recognize, to integrate and express rather than to select ready-made answers. Subjective tests fill the gap which was left open by objective tests and leave the testees approach the problem freely. A balanced testing procedure would include the use of both objective and subjective questions.

3.3.11.2. The Essay Test:

While teaching, teachers always pay special attention to make a kind of balanced teaching to include the productive and the receptive skills necessary to master a foreign language. Of the four skills, writing may be the most difficult skill because it needs knowledge and expressions to convey meanings.

This desired standard of proficiency demands a series of steps to be developed by the learner, including knowledge of vocabulary, spelling, and grammar. Essay tests are probably used frequently because of the great value placed on them, that is, on their need to organize and express ideas in a reliable and accepted form. Satterly (1985) gives us his point of view of essay tests:

> This is probably the most extensively used type of item in Britain outside mathematics and physical sciences. Although it is not much loved by measurement theorists, who have demonstrated its deficiencies over many years, it remain a standard device for the assessment of achievement over a wide range of the curriculum. (P. 114)

Essay types are of great importance to classroom teachers who deal with achievement. Brown (1981) quotes Stecklein (1955) who gives 14 abilities that can be measured by essay items:

1- Comparisons between two or more things.
2- The development and defence of an opinion.

3- Questions of cause and effect.
4- Explanations of meanings.
5- Summarizing of information in a designated area.
6- Analysis.
7- Knowledge of relationships.
8- Illustration of rules, principles, procedures and applications.
9- Application of rules, laws, and principles to new situations
10- Criticisms of the adequacy, relevance, or correctness of a concept, idea, or information.
11- Formulation of new question and problems.
12- Reorganization of new questions and problems.
13- Discrimination between objects, concepts or events.
14- Inferential thinking. (P. 64)

Moreover, teachers can use it because it is easy to construct and develop the habit of relying exclusively on it for its simplicity, but because it is an extremely valid test to measure students' ability to organize their own ideas and express them in a highly effective manner. Brown (1981) gives the following guidelines for writing essay questions:

1- The question should clearly define the task.
2- Indicate the scope and direction of the answer required.
3- Use questions that have correct answers.
4- Allow for "think time".
5- Use more shorter - essays items rather than fewer longer ones.
6- Develop a scoring key before administering the test (PP. 65-67)

3.3.11.3. Scoring the Essay Item:

Although teachers might object here that preparation of a scoring key for each test will take time that they cannot afford, we introduce the two approaches that Satterly (1985) suggests.

1- The analytical method.

> This consists of the production of a check list of points considered essential to a good answer to the question. Deciding in advance which points to look for and the number of marks to be allocated to each point.

2- The impressionistic method.

> This applies where teachers wish, for whatever reason, to award a grade or mark for the essay as a whole. (PP. 120-121)

Scoring the essay items is not easy, so teachers know that they should be careful to maintain the same stand of scoring to all students.

3.3.11.4. Advantages of the Essay Item:

1- Essay tests motivate the students to study and understand concept and principles.

2- They enable the teacher to appraise students' ability to express their ideas effectively.

3- They have the ability to measure complex ideas.

4- The question is very easy to write.

5- They give students freedom of response.

6- Guessing is very limited in this type.

3.3.11.5. Disadvantages of The Essay Item:

1- True and realistic judgement requires much time to be spent on the scoring key.

2- They need competent knowledge of the content area.

3- Their scores are unreliable.

4- They represent a limited sample of contents.

As an interesting end of the discussion about objective and subjective testing items, it is also interesting to quote Ebel (1965) who gives the differences between the two types;the objective and subjective types:

1- An essay question requires the student to plan his own answer and expresses it in his own words. An objective test item requires him to choose among several designated alternatives.
2- An essay test consists of relatively few, more general questions which call for rather extended answers. An objective test ordinarily consists of many rather specific questions requiring only brief answers.
3- Students spend most of their time in thinking and writing when taking an essay test. They spend most of their time reading and thinking when taking an

objective test.

4- The quality of an objective test is determined largely by the skill of the test constructor. The quality of an essay test is determined largely by the skill of the reader of student answers. (PP. 84-91)

Subjective testing, on the other hand, can only be marked by human beings, while the objective type can be marked mechanically. Evaluating a piece of writing like a free composition is virtually an entirely subjective matter. Test reliability may be high in objectively marked tests like true-false tests, but can be low in subjectively marked ones like the essay type.

3.3.12.1. Test Types in Qatar:

In a word, the classroom teacher finds himself faced with having to assess the progress of his students, to judge their readiness for one kind of instruction or another. Depending on the previous assumption, the teacher must decide out of the the whole complex of considerations which has been outlined above what kind of test he wishes to make.

In Qatar, The English Inspectorate gives detailed outline of the proposed objectives to be achieved in a year. Teachers should study these objectives carefully for two reasons. The first is to familiarize themselves with these objectives in

order to do every possible attempt to achieve them, and the second reason is to take these objectives into consideration while preparing to test their students. Teachers use different types of test in the school environment, but teachers who construct these types do not know test classification. That is to say, they do not know that tests of language sub-skills measure the separate components of English such as vocabulary, grammar or pronunciation, while communicative tests, on the other hand, show how well students use the language in actually exchanging ideas and information. Teachers of English do not know whether their tests are proficiency or achievement tests. Their tests are constructed imitatively. We do not want all teachers to be test experts, but to know what type of test they are using and how to construct classroom tests in an appropriate way to benefit from testing.

To sum up, teachers of English in the Qatari schools need to bridge the existing gap between their ideas on testing and what testing really means. This will reinforce their ability to play the necessary role in improving the present situation by applying different types of test. An attempt is made, in this thesis, to indicate some of the options open to teachers on what techniques to use when they want to test their students. Achievement tests that are constructed by teachers provide a closer fit between the tests they construct and the course

contents they are trying to achieve. Each type of test has its own unique characteristics, uses, advantages, and limitations.

CHAPTER 4. VALIDITY AND RELIABILITY-FUNDAMENTAL QUALITIES OF GOOD TESTING:

Topical Outline

4.1 Introduction

4.2 Validity

4.2.1 Literature Survey

4.2.2 Types of Validity

4.2.2.1 Face Validity

4.2.2.2 Content Validity

4.2.2.3 Construct Validity

4.2.2.4 Predictive Validity

4.3 Reliability

4.3.1 Introduction

4.3.2 Definitions of Reliability

4.3.3 Ways of Estimating Reliability

4.3.3.1 Test-retest Technique

4.3.3.2 Alternative Forms of the Original Test

4.3.3.3 Split-Half Technique

Chapter Four

VALIDITY AND RELIABILITY

-FUNDAMENTAL QUALITIES OF GOOD TESTING-

4.1 Introduction

In this chapter, I shall discuss the general considerations that are always important in test evaluation. A test can be regarded as a good test to the extent that it satisfies certain practical and technical requirements.

The most important qualities to consider in the preparation and use of tests are validity and reliability. Both are essential procedures to obtain effective testing. When teachers try to select or construct tests, it is important to ensure that they actually measure what it is intended to measure, yield accurate scores and be relatively easy to administer and score. When I use the word select, I mean the selection of published tests of achievement by classroom teachers. Hopkins and Antes (1985) on the other hand, prefer that teachers should construct tests for their classes rather than select other tests.

> The classroom teacher has a better opportunity to build valid achievement tests for a particular class of students than standardized test makers do because that teacher knows more than any one else about what content has been covered and the behaviours expected. (P. 298)

All good tests possess validity and reliability and other qualities such as practicality which means that the test is appropriate in terms of our objectives and applicable to a certain situation. Validity has been defined by many educationalists when they refer to testing. Now, we turn our attention to what educationalists say about these twin qualities separately.

4.2. Validity:

4.2.1 Literature Survey:

The literature about validity is relatively large, so I shall try to quote the most appropriate ideas to clarify the term and its use in testing in general and in particular in classroom achievement testing.

1-Goodwin and Driscoll (1980) define validity in the following words:

> The validity of a measure is the extent to which it fulfils the purpose for which it is intended. (P. 77).

2-Lado (1977) defines validity by asking a question:

> Does the test measure what it is intended to measure? If it does, it is a valid test, validity is not general, but specific. (P. 30).

3-Adkins (1988) gives the following definition:

> A common definition of test validity is the extent to which the test serves its purpose. (P. 31).

4-Gronlund (1982) expresses nearly the same idea when defining validity:

> Validity is concerned with the extent to which test results serve its intended use. (P. 125)

5-Satterly. (1981)

> Assessments are made for a variety of purposes, the study of validity, is therefore the study of how well those purposes are fulfilled (P. 225).

We are mainly interested here in the meanings of validity with respect to language testing as opposed to the general issues. The concept of validity, as used in testing, can be

clarified by quoting Gronlund (1982) who mentions four points to be noted when dealing with validity.

> 1. Validity refers to the interpretation of test results (not to the test itself).
> 2. Validity is inferred from available evidence (not measured)
> 3. Validity is specific to a particular use (selection, placement, evaluation of learning and so forth).
> 4. Validity is expressed by degree (for example, high, moderate, or low). (P. 126)

Validity serves as an indicator to help us decide whether we are in the right direction or not. If the test measures what we want it to measure we say that the test is valid.

When we deal with validity, we should keep in mind that it refers to the results of a test for a given group of individuals, but it does not refer to the instrument itself. The second thing to be remembered is that validity is a matter of degree. It does not exist on an all-or-none basis. As a result of this, we should avoid thinking of evaluating results as valid or invalid. Validity is best considered in terms of categories that specify degree, such as high, moderate, and low validity. Hopkins (1985) says:

> Tests as used in the educational setting should not be classified as either totally valid or totally invalid but must be conceived as being valid to a particular degree. (P. 298).

4.2.2. Types of Validity:

Validity is traditionally classified into the following types: face validity, content validity, construct validity and predictive validity. These four types are essential for an adequate comprehension of validity and its role in the aim of testing. Each of these basic types will be discussed in turn.

4.2.2.1 Face Validity:

As its name shows, face validity implies that the format of the test reflects the abilities it tests, that is to say, if it looks as though it is testing what it does test, then we say that the test has face validity. It is sometimes considered to have great importance in evaluating a test. Harris (1977) says:

> The most frequently employed type of all is "face validity". Here we mean simply the way the test looks -to the examinees, test administrators, educators and the like. (P. 21).

Goodwin and Driscoll (1980) also give this type of validity the same importance that was given by Harris, moreover, they explain what is meant by face validity:

> Sometimes a measure is said to have face validity. This means that it appears to be appropriate and

> adequate, the format, content, administration and scoring procedures. (P. 81).

When we as teachers construct tests for our students, we are not able to evaluate our own tests. If this constructed test is shown to colleagues or friends, they may discover its ambiguities and criticize it on the basis that it does not look suitable. Heaten (1985) says:

> As constructors of the test, we can become involved in the test that we sometimes fail to stand back and look at the individual test items objectively. Only if the test is examined by other people can some of the absurdities and ambiguities then be discovered. (P. 153).

As a result of reviewing the literature on testing, I found that the educationalists who are interested in testing are divided into two groups as far as face validity is concerned. Some consider face validity as an important type of validity, among them are the following names: Satterly (1985), Goodwin and Driscoll(1980), and Heaten (1985). Others do not consider face validity as a type related to validity, among them are the following: Thorndike (1969, Klausmeier (1971), Hopkins (1985), Hagen (1969), Lindeman (1971), Marshall and Hales (1972), Tuckman (1975), Brown (1980), Swezey (1981) and Doyle (1984). At this stage, it is relevant to quote Gronlund (1981) to represent the second group. He says:

> Face validity should not be considered as a substitute for content validity. In fact, the term face validity is a misnomer; it is not really a type of validity at all.(P. 69).

Finally, it seems to me that content validity can replace face validity and when somebody tries to judge a test, content validity must speak louder than face validity.

4.2.2.2 Content Validity:

When classroom teachers construct tests for their students, they always need to ascertain to what extent the students have learned what teachers have taught them during the school year. What students have studied was presented to them via a systematic curriculum. Teachers need to know how well the items which make up the test match the curriculum objectives and also the actual content of the teaching-learning experiences. In this situation, content validity is necessary to classroom achievement testing. Satterly (1985) speaks of the concept of content validity saying:

> Content validity is an especially important concept for achievement tests and requires that the teacher or test constructor builds into the test not only the topics which were covered, but also items which demand the application of the skills. (P. 227)

Weir (1988) also expresses the same importance of content validity:

> Particular attention must be paid to content validity in an attempt to ensure that the sample of activities to be included in a test is as representative of the target domain as is possible. (P. 25).

Hopkins and Antes (1985) give special priority to content validity when the test constructor is concerned with classroom testing

> For tests developed in the classroom the most important type of validity to be concerned about is content validity. (P. 299)

In the classroom, we teach different skills as well as many aspects of the target language. If teachers want to ascertain that what they construct are valid tests, they can follow a set of useful guidelines for establishing content validity. Weir (1988) quotes Anastasi (1982) who provides these guidelines to be as follows:

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

> under consideration, rather than on the apparent relevance of item content". (P.25).

Content validity is of fundamental importance in achievement tests. When a teacher wants to construct a test, he should examine carefully the content of the unit he is going to test and make sure that the unit-content is faithfully represented in the constructed test. It should be pointed out that classroom teachers can show better content validity than published tests do because published tests constructors are not able to build specificity into a test as well as classroom teachers. Hopkins and Antes (1985) speak in an assertive tone in this respect.

> The classroom teacher has a better opportunity to build valid achievement for a particular class of students than standardized test makers do, because that teacher knows more than anyone else about what content has been covered and the behaviours expected. (P. 298).

Quite obviously from the above, content validity is suitable for achievement testing because the students' achievement is based on a certain plan which the classroom teachers follow to achieve specific objectives during the course of their teaching and the tests that classroom teachers construct should represent these objectives. Lindeman (1971) connects it with achievement testing.

> Content validity is most appropriately considered in connection with achievement testing. (P. 37).

Goodwin and Driscoll (1980) agree with Lindeman when they say:

> The extent to which test items or tasks represent the content and processes of a curriculum universe or domain is known as content validity, most often associated with achievement tests. (P. 78).

It is interesting to say that content validity can be judged before the test is used, provided that we have a clear idea of the curriculum being tested. The contents of the curriculum should be in agreement with the test items being included to test that the students have mastered these contents. Lindeman (1971) speaks of the agreement:

> The overall judgement concerning content validity is based on the extent of agreement between the test and the instructional plan. (P. 37)

In a word, content validity is easily judged by teachers when they construct classroom tests because it has nothing to do with statistical aspects of tests also because it has nothing to do with test scores.

4.2.2.3 Construct Validity:

In addition to content validity which helps us determine how well test scores represent the achievement of certain learning outcomes, we may wish to interpret test scores in terms of some general psychological quality. Construct validity gives us additional understanding of the individual. This type of validity can be well understood if we try to review the various definitions that have been given to construct validity.

1-Karmel (1978)

> Construct validity is ascertained by investigating what traits a test measures, that is, what the test score tells us about a person. (P. 108).

2-Gronlund (1985)

> A construct is a psychological quality that we assume exists in order to explain some aspects of behaviour. Mathematical reasoning is a construct and so are intelligence, creativity, reading comprehension. (P. 72).

Construct validity is connected with the personality of the individual. If testers want to explain the reason that scores differ from high to low or from low to high in different situations, they try to hypothesize personality theories that

might help them to explain what is going on within theindividual. Such knowledge cannot be obtained directly, but through controlled observations of the individual's performance in school. Construct validity is necessary to achievement testing since achievement is a meaningful characteristic of the pupil. This type of validity can serve as an indication of the relationship between what we predict and what test scores show. Gronlund (1981) gives an example about how prediction works in relation to construct validity.

> Let us suppose that we wish to check the claim that a newly constructed test measures intelligence. From what is known about intelligence, we might make the following predictions:
> 1- The raw scores on the test will increase with age.
> 2- The test scores will predict success in school achievement. etc.
> Each of these predictions, and others would then be tested, one by one. If positive results are obtained for each prediction, the combined evidence lends support to the claim that the test measures intelligence. (P. 82).

Dealing with construct validity is a difficult procedure because it demands some statistical analysis. The Encyclopedia of Education (1971) quotes Cronbach and Thomas to indicate this difficulty.

> Statements about construct validity may use various statistics to support interpretations of the test. (P. 170)

Gronlund (1985) also says:

> In general, the process of construct validation involves: (1) identifying and describing, by means of a theoretical framework, the meaning of the construct to be measured. (2) deriving hypotheses. (P. 73).

Construct validity is very necessary to classroom teachers since they deal with achievement in learning. It provides them with additional information about students individually. Sax (1986) gives reason for determining construct validity:

> Determining the validity of a construct is important when the test developer or test user wishes to learn more about the psychological qualities. (P. 295).

Construct validity is of great importance in nearly all types of testing, but we are mainly interested in classroom achievement testing since achievement is a meaningful indicator of every individual pupil. In a word, construct validity aims at identifying the nature and the strength of all factors influencing the students' performance on the intended test. If an achievement test has high construct validity, it should clearly distinguish between students who have achieved high scores at different levels and those who have achieved low scores at the same levels of difficulty.

In language testing, teachers are concerned with motivation, and creativity. Associated with each of these is a

group of highly related behaviours. If teachers want to make sure that their prediction of the ability the students will demonstrate on a given test, they construct a test inspection can be used to determine whether the test has construct validity. If the test has construct validity, scores on the test will vary from student to student the same way that the theory for that construct would predict.

4.2.2.4 **Predictive Validity.**

This is the third type of validity. It can be used for predicting the individuals' future performance or selection. Sax (1980) says:

> Predictive validity coefficients are usually obtained for selection and placement purposes. (P. 297).

Harris (1969) expresses the same idea saying:

> If we use a test of English as a second language to screen university applicants and then correlate test scores with grades made at the end of the first semester, we are attempting to determine the predictive validity. (P. 20).

At the end of the school-year, students always think about the courses they will take or the colleges they will choose the

following year. High school counsellors want to be sure about the present capability of their students. To reach such a decision, they may use tests to obtain the needed information about the individuals' future performance which serves as an indicator for students' selection.

If we want to judge a certain test, we may say that this test is valid and that one is invalid. In this case, our judgment is not correct since the purpose of test validation must be mentioned. A certain test may be valid for one purpose and at the same time invalid for another. Dyer (1965) explains:

> A given test may serve all three purposes (content, construct or predictive) more or less well, or it may have high validity for one purpose and low for another, thus, it is not meaningful to refer to the validity of a test without qualifying the term. (P. 37).

On the basis of classroom testing. As I see it, teachers' prediction is not always valuable and necessary. They need not do that since their task is not prediction, rather it is concerned with teaching in the first place, so they should know more about content validity and how to contruct a content valid test

Now let us examine the classroom tests that are in use in Qatari schools with relation to the previous types of validity.

Teachers, in all levels of learning in Qatar, Primary, Preparatory and Secondary, are not concerned with this quality of good testing. Most of them do not care at all about validity when they construct classroom tests because they know in advance that neither the classroom teacher nor the educational management ever think of analysing test scores for any of the educational purposes of testing. Finally, classroom tests in Qatar are not checked for validity. It is of great value to give an example about real classroom tests in Qatar. At the Secondary level (secondary one) Unit one is called "The World of the Unknown". When one of the teachers who teaches at this stage wanted to test his students, he constructed a test and included items that are not related to the aims and functions of the unit he was testing. The functions of Unit One are the following:

1. Comparing: it's like, it looks like.
2. Suggesting: come on; let's go; let's get out of hear.
3. Asking polite questions: could you...?
4. Expressing disbelief: I don't believe it.
5. Introducing and expressing opinions: as I see it.....
6. Talking about future: I'm going out of here.

When the teacher wanted to test these unit functions, he constructed a test saying in the first item:

"Read the following conversation and then answer the questions below". The teacher introduced a conversation which was taken directly from the textbook and asked the students to find examples of:

1. Comparison: .....
2. Suggestion: .....
3. Advice: .....
4. Obligation: .....

If we compare the functions that the teacher asked the students to identify, we can get the impression that the teacher did not study well Unit One functions. He asked them to find examples of advice and obligation. These two functions were not included in the unit functions listed above. In this situation we say that his test is not content valid since it did not respect the rules of content validity. The degree of content validity is determined by making a comparison between the content of the test with the content of classroom instruction to test how well the instructed materials have been achieved by the students during a predetermined length of time. The adequacy of sampling is also important here. That is to say, the test should represent the taught materials. In the previous example, the teacher tried to test only two functions out of six that were included in the unit functions above. Thus, the teacher did not respect the adequacy of sampling in his test.

4.3 Reliability:

4.3.1 Introduction:

Those who are interested in testing give validity and reliability primary consideration when evaluating a test. This quality of good testing gives the test user the extent of how far he can depend on the test results as a real indication of the ability being measured by the test.

Basically, this section attempts to deal with the problem of the consistency of testing. Readers are then introduced to the meaning of reliability through reviewing different opinions about this essential term.

4.3.2 Definitions of Reliability:

An important aim in the development and administration of tests is to identify potential sources of error in a given measurement of language ability and to minimize the effect of these factors on that measurement. The classromm teacher sometimes asks, "How similar would each student's scores be on the first test and a subsequent test if the same tests were used?" The answer to this question is based on the test reliability. The following definitions of reliability will concentrate on the stability of the test scores.

1-Lindeman (1971)

> Reliability is defined as the consistency with which a test measures whatever it measures. (P. 43)

2-Tuckman (1975)

> Reliability indicates the degree to which a test is consistent in measuring whatever it does measure. The degree to which the test measures the same thing time after time. (P. 254)

3-Harris (1977)

> By reliability is meant the stability of test scores. A test cannot measure anything well unless it measures consistently. (P. 14)

4-Gronlund (1985)

> Reliability refers to the consistency of measurement, that is how consistent test scores or other evaluation results are from one measurement to another. (P. 86)

All the previously mentioned quotations concentrate on one thing, which is the consistency of test scores with which test performances are evaluated. I am mainly interested in the consistency of the scores because they are affected by a number of factors, chief among them being the adequacy of the sampling

of the tasks being tested. The second factor that affects reliability is the difference from one administration to another. That is, different markers give different scores to the same test. Finally, the purpose of the test also affects reliability on the basis that a certain test may be reliable for one purpose and unreliable for another. These are not all the factors that affect reliability, but the most important ones. The factor that affects the scores has been looked at differently by testers. Hopkins (1969) relates it to the individual, while Sax (1980) relates it to the item itself. Hopkins (1969) says:

> When we have extracted two scores from a single testing, both scores necessarily represent the individual as he is at the same moment of time. Every event lasting only a few minutes will affect both scores. (P. 184)

Sax (1980) puts it in another way:

> The number of items on a test also affects reliability. The more items there are, the higher reliability will be. (P. 270)

It seems to me that both the individual's physical mood such as illness or recovering from illness or even his psychological characteristics at the time of the test such as motivation or fatigue, and the test item quality such as

objectivity and difficulty affect the test scores. I conclude that validity and reliability are of great importance when test evaluation is taken into account. Both of them are needed and each one is necessary to the other

4.3.3 Ways of Estimating Reliability:

Reliability, as we have seen in the definitions, may be obtained by one of several procedures. The choice depends on the use of the test scores. Although there are various methods of estimating reliability, the most commonly used are those that have been mentioned by Karmel (1978) to be as follows:

1- Retest subjects with the same test.
2- Alternate form of the original test.
3- Split-half which involves a division of the test into two parts. (P. 112)

Now, I shall deal with these three ways separately.

4.3.3.1 Test-retest Technique.

This is the simplest way to estimate reliability. In this way, we test the same students with exactly the same test. The obtained scores from the same individuals are expected to be

the same. Satterly (1985) considers this technique as the easiest to administer.

> This is probably one of the easiest to obtain since it involves only two administration of the same test to a given group of children. (P. 196)

Karmel (1978) gives an interesting example to explain this technique.

> If a physician, for example, wanted to check on the accuracy of a nurse's ability to measure patient's weight and height. The physician might ask the same nurse to measure each patient twice, using the same procedure. (P. 112)

While repeating the same test twice, particular attention should be paid to the time interval that separates the two administrations. Quite obviously, the obtained scores will sometimes vary between the two administrations. If the students have the chance of two or three days interval between the two tests, they will often get either the same scores or the second will be better than the first. This is because the students may avoid their previous mistakes on the second test. Whatever the interval, it has advantages and disadvantages. Gronlund (1985) says:

> One important factor to keep in mind when interpreting measures of stability is the time

> interval between tests. If this time interval is short, say a day or two, the consistency of the results will be inflated, because pupils will remember some of their answers from the first test to the second. If the time interval is long, say about a year, the results will be influenced not only by the instability of the testing procedure, but also by actual changes in the pupil. (P. 90)

The scores variation depends more or less on the time interval which has its clear effect even on the individuals themselves. Some of them may lose interest in taking the test again. Lado (1977) explains this point saying:

> Some students may lose interest when asked to repeat a test, while others may not, they may be occupying different seats on the retest and this may affect their performance. There may be differences in noise other disturbing elements at the time of the two administrations. (PP. 333-334)

The final word on this technique is that it has nothing to do with classroom testing. It is of great importance to constructors of standardized tests when they want to ascertain the reliability of their tests.

4.3.3.2 Alternate Forms of the Original Test:

The second way of estimating reliability of a test is the alternate forms of the same test. In this technique two different forms of the same test are used. These two forms

should be equivalent in length, difficulty, time limits and format. Other aspects may also be taken into consideration such as the area of knowledge on which the two tests will depend. Karmel (1978) puts a great emphasis on test-items:

> Each form should contain the same number of items covering the same kind of content and arranged in the same format. All aspects of the test including the degree of content. difficulty, instructions, time limits and so forth - must be comparable. (P. 114)

Our time interval is still existing even in this technique, but it is recommended to be a short time between the two forms administrations. Gronlund (1985) suggests that:

> The two forms of the test are administered to the same group of pupils in close succession (P. 91)

Thorndike (1964) agrees with Gronlund saying:

> They may follow each other immediately if we are not interested in stability over time. (P. 182)

This technique is usually used for school-wide standardized tests.

4.3.3.3 Split-half Technique:

In this technique, the same test is divided into halves for scoring after a single administration. In this situation, two scores will be obtained for each individual student. The two halves will be scored as Gronlund (1985) suggests:

> To split the test into halves that are most equivalent, the usual procedure is to score the even-numbered and the odd-numbered items separately. (P. 92)

During the time of administration, we cannot be sure that the individual student was in the same mood in the two parts of the same test. This means that a kind of limitation should be taken into consideration. At this stage, it is necessary to give an answer to those who may ask "How can we make tests more reliable. The answer is given by Hughes (1990) through many suggested procedures:

1. Take enough samples of behaviour.
2. Do not allow candidates much freedom.
3. Write unambiguous items.
4. Provide clear and explicit instructions.
5. Candidates should be familiar with format and testing techniques.

6. Use items that permit scoring which is as objective as possible.
7. Provide a detailed scoring key. (PP. 36-41)

4.4 Relationship Between Reliability and Validity:

The most important quality, test users must take into consideration, is validity. Reliability is necessary for validity in the sense that test scores that are not reliable cannot provide a basis for valid interpretation and use of language tests. Reliability is very important procedure to obtain validity. Gronlund (1982) says:

> Reliability is a necessary, but not a sufficient, condition for validity. (P. 33)

During the test construction procedure, when a teacher tries to choose the sample of the learning objectives to be included in his test means that he is on the right way to establish content validity. On the other hand, when adequate numbers of items are included in the sampling this also means that he is moving toward establishing reliability. When we increase the reliability of our measures, we also satisfying a necessary condition for validity; in order for a test score to

be valid, it must be reliable. Most discussions of these two important qualities of good testing speak about the differences between these two qualities, rather than their similarities. I believe that both of them can be better understood by saying that they are complementary aspects of estimating sources of variance in test scores. Bachman (1990) speaks about the difference between these two qualities:

> In estimating reliability we are concerned primarily with examining variance in test scores themselves. In validation, on the other hand, we must consider other sources of variance, and must utilize a theory of abilities to identify these sources. That is, in order to examine validity, we need a theory that specifies the language abilities that we hypothesize will affect test performance. The process of validation thus must look beyond reliability and examine the relationship between test performance and factors outside the test itself. (P. 239)

The two qualities are always of great importance to those who construct tests, but if a kind of choice is to be made, validity is more important to be achieved by the tester.

In Qatar, there is no sign that either validity or reliability are considered. That is to say, teachers do not care at all about these two essential qualities of a good test. These two qualities were dealt with only on experimental basis when the Ministry of Education asks a testing expert to evaluate classroom tests. Teachers of English do not care about these two qualities because they do not know how to deal with

them. They think that any set of questions can be called a test.

Chapter Five

CASE STUDY BASED ON "THE CRESCENT ENGLISH COURSE"

Topical Outline:

5.0 Introduction. (course background)

5.1 Course objectives.

5.2 Analysis of three teacher-produced tests.

5.3 Test one.

5.3.1 Testees' background.

5.3.2 Test specification.

5.4 Test two.

5.4.1 Testees' background.

5.4.2 Test specification.

5.5 Test three.

5.5.1 Testees' background.

5.5.2. Test specification.

5.6. Implication for future practice.

Chapter Five

CASE STUDY -BASED ON "THE CRESCENT ENGLISH COURSE"

5.0 Introduction: (course background)

In this chapter, the survey is meant to identify present conditions to provide information on which I can base sound recommendations. It is an exploratory chapter in nature, with emphasis on breadth rather than deep discussions. I am concerned here with determining the current status of testing in the "Crescent English Course".

As pointed out in chapter one, in Qatar we are teaching and testing "The Crescent English Course" which is based on the communicative approach. The Ministry of Education aims at producing bilingual students able to interact effectively with other English-speaking communities. Taking this into consideration, the Ministry of Education decided to put the communicative approach into action to provide the students not only with formal knowledge of the target language, but also with the ability to communicate and use English effectively. The course is based on the current understanding of the communicative nature of language. Lessons are designed, as far as possible, around activities which are motivating in themselves and to some extent reflect the interests of the

students to introduce them to the social use of language through the well defioeD actirity of group work and pair work.

The materials were produced by Oxford University Press (OUP), eight-level Crescent English Course, a Communicatively-based series that was specially written for Arab learners by a British team. Appendices from 4 to 10 show the materials. The organization of the course varies in length with each level. The teacher's guide, however, not only breaks the materials into units but also into steps (see appendix 11).

The recorded materials play an important part also in introducing the students to the target language spoken by native speakers of English. The course is well organized, progressing from listening and speaking to reading and writing. Repetition, recycling and revision of the taught materials are well planned and built into the course. There are 188 schools in Qatar in addition to another 4 specialized schools, Religious (2), Commercial (1), Industrial (1).

5.1 Course Objectives:

When I use the word "objectives", I mean what the students should be able to do after a prescribed time of instruction. The Crescent Course was criticized on the basis that it was put into action without clearly prepared objectives. The objectives of the "Crescent Course" should have been set in advance. Qotbah (1990) says:

> What is strange about the "Crescent Course" is that it was written without clear objectives in mind. (P. 31)

The objectives should be set in advance to help teachers to try to achieve them. If teachers do not know what they are aiming at, they will not be able to decide whether they have reached the destination or not. I could not establish a definite date of the production of the Crescent Course objectives. There is a contradictory specification of date as far as the objectives are concerned. Dr. Bratton, English Language Consultant in Qatar, says in a report which was presented in a seminar in Bahrain 12-15, November, 1983:

> The aims and objectives of English language in Qatar are set down in a document prepared in June 1983 by a sub-committee of the ELT Supervisory Committee. (P. 91)

While Qotbah (1990) says:

> In 1985 English teaching objectives were formally produced by the Ministry of Education. (P. 32)

The objectives that have been set by the the Ministry of Education for the three levels of learning (Primary, Preparatory and Secondary) are listed in appendices 1, 2, and 3.

As pointed out in chapter one, tests in Qatar are carried out at several parts each term. The first two monthly tests and the mid-term are constructed by classroom teachers. At the end of term, pupils are given tests that are constructed by the Inspectorate to reflect the aims of the Crescent English Course. These tests also serve the purpose of allowing comparison between students from different schools and setting national standards for all schools.

In principle both kinds of test are introduced to fulfil the same purpose, that is, they are tests of achievement. Similarly they should have the same kind of validity and reliability.

5.2 Analysis of Three Teacher-Made Tests:

As there are three stages of learning we decided to focus on one test from each stage. The tests were selected in the following way. A collection of tests has been established over ten years from schools and colleagues. From this collection, three have been selected from among examples of weak tests. We can say here, however, that although there are an increasing number of good tests, this is due mainly to imitation of tests used by other teachers, and the use of the sample tests provided by the textbook. Our purpose in selecting these examples of weak tests is primarily to demonstrate the depth of the problem if teachers are not helped with test-construction. Classroom tests are generally prepared, administered and scored by one teacher. In this situation, tests should be based directly on course objectives.

As we have seen in chapter three, language tests have many purposes and quite often the same test may be used for two or more related purposes. In our analysis, the following procedure will be followed. The test deficiencies will be pointed out as compared to what should be done in the same situation.

Classroom tests constructed to measure achievement are made up of tasks based on the content covered during a period of time. The problem of the testmaker is to select from that very large number a sample which best represent the taught

materials. The test is best considered as a sample if it is composed of set of items chosen from a large number of activities that have been taught to the students.

The best way to state our case is to give actual examples of teacher-produced tests. Let us call them "Test One, Two, and Three". Test One is a typical example of Primary level, two represents Preparatory level, and three stands for Secondary level.

5.3. Test One: (see appendix 12)

5.3.1 Testees' Background:

The testees are young learners of English from Qatar and many other Arab countries. They are between 10-11 years old. They are in their second year of learning English (Primary 6). They wish to use English for a number of purposes in their country and to make outside contacts when they grow up.

5.3.2. Test Specifications:

A paper-and-pencil test used to measure achievement should be made of tasks based on the content covered in the classroom. These tasks should relate to what students have studied during a limited period of time. Testing should be understood in a way that it is intended to make a change for the better as noted in chapter three.

In this analysis, I shall try to apply the ideas discussed in chapters two and three on each of the selected tests. When a teacher wants to test his students, he should have a clear and predetermined aim for his test. To achieve this aim or aims, he must prepare test specifications and then pick out the most promising areas for testing the well known skills in learning a

foreign language, that is, listening, speaking, reading and writing. If we look at appendix A, we can find that the test tries to test the following areas:

1. Identification.
2. Time.
3. Completion.
4. Opposites.
5. Numbers.
6. Description of locations.
7. Description of people.
8. Writing a reply to an invitation.

The specifications that the test is trying to measure should be selected carefully from a larger domain. A classroom test should be developed in a way to provide feedback which will serve as a reliable indicator to inform students and parents about student progress. To clarify the point, let us list the predetermined objectives of the Unit the test is trying to measure. The Unit aims to teach:

1. Parts of the body.
2. Colours.
3. Possessive adjective.
4. Questions about identity.

5. Questions about locations.

6. Plurals.

7. The recognition and writing of the capital alphabet.

8. Numbers: 1-12.

9. To teach: reasons, directions, Past Simple, expressions of time.

In this test, we can find the following shortcomings:

1. The test constructor did not take the unit objectives into consideration, although the obectives are sufficient to generate balanced test activities to measure pupils' abilities that have been acquired during classroom instruction. As we have seen in chapter three, the achievement test should be used to monitor the learning process and to cover a limited segment of instruction such as a unit and attempts to measure all important outcomes of that unit.

2. Question No. 2 is mainly written to test the pupils' mastery of numbers 1-12. If we have a look at the question, we will find that the numbers have already been given in the instructions. What is required from the pupils is just to write the word "O'clock". If the teacher wants to test the pupils' ability to write the word "O'clock", he can test it in one item but not in three consecutive items.

3. To test the Past Simple and the Present Continuous, the teacher tried to test these two tenses in question No. 3, but failed to design the appropriate question. The teacher

asked the pupils to complete three sentences;

1. We are playing in the ....
2. Yesterday, I ate a....
3. We.....to school.

In the first sentence, the Present Continuous has already been used in the sentence. We can ask what is the teacher going to test? The sentence should have been introduced in the following way. "We .....(play) in the garden now". The teacher can, then, make sure whether the pupils know how to form the Present Continuous or not.

The test lacks content validity because the emphasis is not placed on measuring mastery of learning tasks that have been introduced in the unit objectives and it does not provide feedback to the pupils about specific learning errors in areas where they have not yet achieved mastery. Furthermore, the test does not make sure that all the major aspects of the instructed materials are covered by the test items. As we have pointed out in chapter three, the purpose of classroom testing is to allow the teacher to achieve certain aims after a completion of a part of the learning course. In this test, the teacher will not be able to plan remedial teaching to overcome deficiencies pointed up in the testing procedure that has been applied in the classroom.

The test also lacks reliability because it does not give adequate sampling of the taught materials. As we have seen in

chapter four, the more samples of students' performance we take, the more reliable will be our assessment of pupils' knowledge and ability.

5.4 Test Two (see appendix 13)

5.4.1 Testees' Background:

The testees are young learners in their 5th year of learning English. They are between 14-15 years old. They wish to learn English for general purposes.

5.4.2 Test Specification:

This test is supposed to be an achievement test which must be a systematic procedure for measuring the extent to which a person has acquired certain information or mastered certain skills as a result of specific instruction as noted in Chapter Three.

The selected test tries to measure only two skills. That is, reading and writing. It does not faithfully reflect the objectives that have been taught throughout the school year. In the real classroom situation, the teacher did not teach only reading and writing but new vocabulary and grammar as well. Thus, the test lacks content validity. Reliability is also affected by the inadequacy of sampling of tasks. Byrd (1986) emphasizes the idea that testing in Qatar is restricted to reading and writing:

> Testing generally received a separate sort of emphasis. While the day-to-day learning materials generally emphasized the four skills and often were engaging and stimulating, the testing procedures--particularly important examinations--were the opposite. Usually restricted to reading and writing. (P. 21)

5.5 Test Three (see appendix 14)

5.5.1 Testees' Background:

The testees are young learners from Qatar and other Arab countries. they are in their 6th year of learning English. They are between 16-18 years old. They hope to use English for general purposes in their future life.

5.5.2 Test Specifications:

Classroom tests can be used for a variety of purposes as noted in chapter three. This test should have been designed to monitor pupils' learning progress and to provide feedback to pupils and teachers. It was given during instruction to decide whether pupils possess the prerequisite skills needed to succeed in part of the course, to reinforce successful learning and reveal learning weaknesses in need of correction. The test did not fulfil the previously mentioned purposes because the teacher did not take the materials he has taught to his pupils into consideration when he constructed the test. The test tries to measure two units, one and three, two is omitted. Let us list the objectives these units are supposed to teach:

Unit One:

Functions:

1. Comparing: It is like, it looks like.
2. Suggesting: Come on; let us go; let us get out of here.
3. Asking polite questions: Could you...?; Would you mind...?
4. Expressing disbelief: I do not believe it.
5. Introducing and expressing opinions: I am convinced; there is no doubt in my mind; as I see it.
6. Talking about the future.

Grammar:

1. Referring pronouns to their subjects.
2. Past Continuous and the Simple Past.
3. Direct and reported speech.

Unit Three:

Functions:

1. Suggesting: How about...; you should...; you ought to...; if I were you...; we'd better...;don't you think you might...; the best thing for you to do is...; I'd prefer to...

Grammar:

1. Question forms (including question intonation)
2. First / third person transformation.
3. Direct / reported speech.
4. Present Perfect.

Test Fairness: This quality means that each student should have an equal chance to demonstrate his knowledge. As far as test three is concerned, the test lacks this essential quality because it concentrates on the writing skill, which is the most difficult skill, more than the other skills (e.g. No. 1-2B-3-4 & 5). These questions require the student to express himself in his own words using information from what has been instructed to him and from his own background and knowledge but not factual information. The expected result will not be satisfactory because the teaching process itself and the classroom teacher do not encourage pupils to organize their own ideas to express them effectively. On the contrary, the classroom teacher usually dictates ready-made answers that he thinks may be included in the test. This device retards the

students' role in organizing their ideas.

If we make a comparison between what the test includes and the objectives of the two units, we can find out that the test lacks both content validity and reliability because it did not test what it is supposed to test. As we have seen in chapter four, teachers need to ascertain to what extent their students have learnt what they have taught them during the school year. How can the constructor of this test reach a decision that his students have mastered the predetermined objectives? Content validity has become a problem here because of the difficulty involved in defining the areas of the course from which the sample is to be selected. The classroom teacher did not make a systematic analysis of the domain to make sure that most of the major aspects are covered by the test items in correct proportion. The test did not include either reading or writing. It is just quick jumps from one idea to another, in contrast to test two which concentrated only on reading and writing.

In Qatar, school examinations are not really a reliable indicator of the students' actual achievements because the present testing procedures do not make a reasonable balance in testing the four skills. The three selected teacher-produced tests give the impression that there is no relationship between teaching and testing. These tests should be based directly on a detailed specification of the course objectives. As noted in

chapter three, classroom tests should make an adequate sampling of the taught materials.

The secondary school examinations should test the development of competence through the teacher's effort to select the appropriate items that may serve the purpose of testing in this advanced level of learning. The difficulty of the test items to be included in the classroom test depends largely on whether the test is being designed to describe the specific learning tasks to be measured to make sure that the students can perform (Criterion-Referenced) test or to rank students in order of their achievement (Norm-Referenced) test. These three selected tests may be classified as Criterion-referenced tests because the classroom teacher is more interested in learning tasks and the students mastery of these tasks rather than ranking them in order of their achievement. This is what really happens in the present testing situation in Qatar.

In chapter one, we have indicated that testing in Qatar plays an important role only in deciding the students' future position. We have also pointed out that testing in Qatar is an objective in itself more than a means of improving learning. These three selected tests indicate that the assumption is true. There is no need to repeat what we have said earlier about the problems of testing in Qatar.

5.6 Implications For Future Practice:

It is being increasingly realized that just testing does not mean anything unless we make use of the test results. This procedure will certainly help in taking the students to a better stage of learning. Teachers, on the other hand, can benefit from having detailed understanding of the individual pupils who they are currently teaching.

There are many shortcomings in the school examinations that are currently in use in Qatar. Testing in Qatar can be looked at as an objective in itself. If we want to examine the present problems of testing in Qatar, we can diagnose the following:

1-The idea of centralization has one advantage and one disadvantage. When the English-Inspectorate members construct school examinations, they take into consideration only the objectives of each class. This is commonly considered a good and important step in test construction, but not the only one. They can hardly judge the areas of concentration while the teaching process is in action. It is of great importance to state here that the idea of centralizing school examinations has at least one advantage which is setting one standard to all pupils taking the test.

2-The second major problem in the testing domain is the lack of concentration on the importance of vocabulary. Needless to say, vocabulary is the basis of any sentence construction. It is a fact that a skilled reader recognizes words of the target language as a whole. He does not have to examine the word letter by letter. This is so in the case of familiar words. The same applies, of course to the recognition of strings of words, such as "once upon a time", "on the other hand" etc. Although the previous idea is more related to teaching than testing, it is needed in the testing procedures. The justification for this idea is that we, as teachers, always face a problem while administering school examinations. Pupils always ask questions such as" How can we answer this item ? What is meant by this question ? ". If the pupils have the ability to recognize the word-formation, they can easily read them and if they can read correctly, then there will not be any problem while answering the test items. Another fact, which leads to bad results, is that teachers should train their pupils now and then on how to answer items of different kinds. If the test instructions are continuously repeated, they will be familiar to the pupils. This problem should be dealt with at the beginning of Primary (5) where the pupils begin to learn the target language. Instead of solving the problem at this stage, what happens is that things are done by halves. We must not lead a hand

to mouth existence without taking the future into consideration.

The use of the Arabic in the examinations' instructions retards the students' progress in understanding what their teachers want them to do, When they are given the instructions in the target language, they always either complain or ask for translation. This is a result of the lack of practice about how to answer a given item. So teachers ought to make sure that their pupils are familiar with question headings. It is necessary to admit that the English Language-Inspectorate members do their best to supply teachers, at all levels of teaching, with the needed question headings at the beginning of each school year. When this is done, we can give our pupils the opportunity to achieve comprehension and help them pass to the production stage through enough practice to use the test instruction in the target language.

3-The idea of testing is not well understood by most teachers. They think that the role of testing is finished when they finish scoring the test. This is the obvious reason behind testing. The idea that testing helps teachers, through appropriate instruments, to set realistic standards of achievement for group or individuals, is not yet realized by most teachers.

4-The present examinations lack readability. That is to say, these examinations are useless if students cannot read them. The examinations must be written in the students' reading level, not in that of teachers.

5-During the test-time, a high percentage of the students' time is sometimes spent on routine (non-work) instructions. For example, some teachers may interrupt the students while answering the questions saying: "attention please, when you finish, please go to room No. 4 to have your oral test" or "Please try to come on time tomorrow" This means that the students are always disturbed by the school personnel who are in charge of tests' administration. The classroom, during the test, should not be dominated by them, instead it should be work-centred classroom to provide the pupils with the needed atmosphere to think quietly.

"6-Although the oral test is the best opportunity for the classroom teacher to test the pupils' ability to use the target language and where the teacher has a unique chance for examining the pupils' mastery of the strategies of negotiation, classroom teachers do not benefit from this opportunity because they always consider that bad scores on an oral test leave irreparable damage to the testee's final standing. Taking this into consideration, teachers give the

testee endless chances to improve his situation. As a result of this situation, there is always a gap between the testee's marks on the written examination and the oral. It is true that experienced teachers use neither criteria nor a rating sheet and score according to their past experience, but as far as the situation in the Qatari schools is concerned, not all teachers have much experience. In general, or at least according to my understanding, no confidence can be placed either in the oral testing procedures or in the test's results.

Chapter Six

CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

In this chapter, I shall try to demonstrate how my earlier chapters and my investigation of the problems of testing in Qatar have led me to the following conclusions. The chapter will be of two parts; the first will summarise the earlier chapters and will also concentrate on the conclusions I have reached as a result of investigations and analysis of the present state of testing in Qatar. The second part will be suggestions about what needs to be done about the present situation, based on my earlier review and analysis of testing.

When we talk about testing in the teaching-learning process, we are really talking about a serious problem that needs urgent solution. I think that testing is still problematic in Qatar. The absence of good testing is one of the most important reasons for the low standards of achievement in Qatari schools. The present testing procedures are unable to achieve any educational goals.

As indicated in chapter one, the main purpose of the thesis is to survey the present system of testing in Qatar and relate it to the development of testing in general in order to place testing in Qatar in the appropriate stage of development. Another purpose, the study is trying to achieve, is to describe the relation between testing and teaching in Qatar. It was also stated that it would be of benefit to propose ways of modifications for improvement.

Chapter 2 deals with the history and development of language testing in general. This will also enable us to justify our ideas about the needed modification. It also relates, in particular, the stages of development of testing. The aim is to be able to know our position in this development. With regard to the Qatari testing state, the time is highly suitable to dismiss traditional techniques of testing and apply modern techniques.

As we have seen in chapter two, testing should play an important role in teaching to help both the teacher and the student at the same time by measuring the progress that has been made to achieve the predetermined objectives and sometimes to modify the ways of presentation. Classroom tests should be analysed by educational personnel to overcome present problems or to reduce the weaknesses and reinforce the strengths of the teaching-learning process. Existing tests can never diagnose the shortcomings of the textbooks being taught or the teaching

procedures that teachers adopt.

Chapter 3 offers an in-depth look into the significance of testing in the classroom and its purposes. It also deals with the burdensome responsibility of teachers to construct their own classroom tests. Objective and subjective tests are thoroughly dealt with.

Chapter 4 provides a description of the two essential qualities of good testing, validity and reliability and the close relationship between these qualities.

Chapter 5 provides a discussion of what is going on in Qatar as far as testing English is concerned. This discussion is based on specific teacher-produced tests to point out the deficiency of classroom test construction.

In this chapter a summary of the principal ideas of the whole thesis will be given in order to help finalize the research conclusions and to suggest the appropriate recommendations to remedy the situation.

6.2 Conclusions

The conclusions may include the following:

1. It was observed in chapter One that there was a shift in the Qatari situation in the form of a change from the Structural Approach to the Communicative Approach to overcome the deficiencies of the the previous approach which neglected to a great extent the listening and speaking skills. Unfortunately, the new adopted approach still neglects listening and speaking while teaching and testing. Testing the students' oral production is also still traditional. I mean to say that although the oral test is one good opportunity for the students to apply what they have learnt to express their ideas, most teachers use stereotyped questions at all levels.
2. The Structural Approach in teaching and testing was abandoned because it used to concentrate on asking the students to memorize list of words to pass the exam. In the new adopted approach, Communicative Approach, the idea of memorization has been changed from memorizing lists of words to include short passages from the curriculum to pass the test. That is, there is less concentration on the skill of writing.

3. Teachers' preparation, by the Ministry of Education, is not enough neither in teaching nor in testing. They should be adequately prepared for their future task. The training courses in Qatar failed to achieve the intended results because the time was not enough to familiarize teachers with textbooks, teaching and testing procedures the teachers must follow while dealing with the Communicative Approach.

4. The learners lack adequate motivation to learn a foreign language, a fact which was clearly identified by both testing experts and researchers.

5. The testing procedures that are in use in Qatari schools are still traditional in the sense that they are not analysed to point out weaknesses or strengths in students' achievement, but to pass students to a higher stage of learning.

6. The important role of teachers in the testing domain is ignored. This may make them negative in improving their knowledge on testing. It means that the English Language Inspectorate members do not trust teacher-made tests.

7. In the present testing process, no notice is taken of the importance of the reliability and validity of the tests being used in the classroom.

6.3. Recommendations and Suggestions:

Recommendation 1 :

There should be more concentration on the skills of listening and speaking when testing students' achievement:

It may seem rather odd to test listening separately from speaking, since the two skills are typically exercised together in oral interaction in testing, but in real life situations there are occasions, such as listening to the radio, listening to lectures, or listening to railway station announcement, when no speaking is called for. Because of the emphasis today on direct communication in a foreign language, the skill of listening has become the object of growing attention. In the teaching process, students must rely on their ears to understand what the teacher is saying in the class in order to achieve overall comprehension. While testing the skill of listening, the main objective of a listening test is to evaluate comprehension. In learning English as a foreign language, we can not and must not ignore the listening skill because we sometimes use taped radio broadcasts and recorded passages from the textbooks to test the students' ability to achieve overall comprehension. Valette (1977) gives an

example to explain the importance of overall comprehension in learning a foreign language:

> A person learning English as a second language and who is familiar with vocabulary and structure but who possesses only the most rudimentary ideas about pronunciation will easily be able to distinguish between taking a bus and taking a taxi since bus and taxi sound considerably different. But this same foreigner may not be able to understand the difference between "Dad is washing the dog outside" and "Dad is watching the dog outside" P. 15

Like writing, speaking is a complex skill requiring the simultaneous use of a number of different abilities which often develop at different rates. Teachers should not ask stereotyped questions which are supposed to be the beginning of the oral interview. The interview should begin with social questions such as: "How are you today?" "What city do you come from?" "How long have you studied English?" Such questions, at the beginning of the interview, will serve the double purpose of helping to put the candidate at ease and how well can he function in social situations. What happens in Qatar is that teachers often ask only the social questions at all levels.

Recommendation 2.

There should be more concentration on the skill of writing:

There is a noticeable lack of emphasis on the writing skill both in teaching and testing. The absence of such emphasis has resulted in incorrect presentation of sentences on any writing test. Writing tests must consequently be structured to measure the various aspects of students' progress toward the acquisition of this difficult skill. We have to set writing tasks that are properly representative of the population of the tasks that we expect the students to be able to perform. Some teachers are not aware of importance of writing. Most of them think that the Communicative Approach stresses the idea of oral production apart from writing.

Recommendation 3.

Teachers' of English should be adequately prepared in testing:

Teachers' preparation is not enough in the present situation in Qatar. They should be well prepared on how to test the language they teach. The existing training courses failed to achieve the intended results to

familiarize teachers with the most recent procedures on testing English as a foreign language because the inspectors who are responsible for training the teachers are not well qualified in testing. Personal experience is not enough in this respect. The Ministry of Education has appointed Mr. Roger Nunn, native speaker, as a testing coordinator but he has been recently transferred to Qatar University to teach English at the English Language Teaching Unit (ELTU). Teachers of English in Qatar should be adequately prepared for their future task not only in teaching the communicative approach, which Qatar adopts, but also in testing this approach effectively. The three teacher-produced tests provided in appendices 1. 2, and 3 reveal the teachers' lack of knowledge about test-item construction.

Recommendation 4.

Testing should act as an identification of the students' lack of motivation:

The learners lack the adequate motivation to learn English. It is the teachers' responsibility to raise this low motivation through appropriate procedures of testing.

This low motivation is not an assumption but a fact which was clearly identified by testing experts and researchers alike.

Recommendation 5.

Teachers should analyse school-tests to diagnose weaknesses to provide basis for remedial teaching:

The purpose of classroom testing should be to diagnose the weaknesses of the teaching-learning process in order to devise suitable remedial teaching. Teachers who do not test to diagnose their ways of presentation cannot do their job properly.

Recommendation 6.

Teachers' involvement in assessing the learning process is a prerequisite:

In the future, Clearly teacher involvement in assessing the learning process is of urgent need to realize the newly applied ideas in the testing domain. This involvement will maximize the opportunity for teachers to experiment their ideas and gain experience on improving

the quality and effectiveness of their testing techniques. Teachers should be given more freedom, after their preparation in testing, to test their students because they live the classroom life in its real meaning.

Recommendation 7.

Clear indication should be given to the importance of reliability and validity of classroom tests:

Most treatments of the subject put validity first, as being the most important quality of a good test. Reliability, on the other hand, generally affects validity and validity cannot be fully appreciated without a basic understanding of reliability. Without these two qualities, a test would be a poor investment in time and money. Teachers should certainly understand what these concepts mean and how to apply them.

The present testing in Qatar concentrates on measuring the students' power of recall rather than the basic skills. The result of such procedure is very clear when we find some students who know the meaning of many English words, but they fail to communicate with others fluently. If we want to evaluate the present testing procedures, we can say that there

is no agreement between teaching and testing. Learning materials generally emphasize the four skills, while testing is the opposite, usually restricted to reading and writing. The writing skill is no more than ticking off, circling the correct answer, filling in blanks, or answering questions. Thus the present testing procedures fail to focus on writing in the real meaning of the word and do not treat the oral or listening skills. Tests as they are now, bear little relationship to what is happening in the English language curriculum.

Present testing in Qatar is not able to cope with the current communicative climate. It is axiomatic that test tasks should as far as possible reflect realistic situations. They should cover as wide as possible a range of the activities the students have learnt during their school year. Despite our best efforts to provide clear instructions and precise task specifications, test takers do not always respond in ways we expect. The way we test our students is not the ideal one because what we present to them does not reflect what they have been taught in the class and our test format does not cover as wide as possible the activities we taught in our classes.

The problem we face in Qatar is that teachers of English often try to simplify the textbook through, what teachers think to be, additional and remedial exercises to reinforce the strengths and reduce the weaknesses in the students' productive ability. What really happens is exactly the opposite. The

students depend completely on the provided exercises which will damage the students' ability to learn the target language. The clear result of this way of teaching is that when we test our students, we test their memory production rather than their language competence.

The role of testing in teaching has many purposes and it should be used to direct the learning-teaching process not only for the benefit of the pupils and teachers, but those outside the classroom such as the Ministry of Education, other schools and universities that depend upon reliable knowledge of pupils' achievement.

There is a gap between teachers, on the one hand, and the Inspectorate on the other. We can explain this in the sense that there is no cooperation between teachers and the Inspectorate. The Inspectorate members construct some sample tests and distribute them to teachers when they visit their schools. Distributing these papers to the teachers is the goal of their construction and nothing else. Inspectors never ask about the feedback of tests that are being applied in the classroom by teachers. We suggest that when the Inspectorate members construct tests, they supervise the application of the suggested procedures and ask classroom teachers about the expected interpretation of tests results in order to achieve the highest possible effects of these tests. The second important procedure that should be taken is to train

inexperienced teachers on how to construct their tests to evaluate what they teach in the classroom. There is a teacher training course for inexperienced teachers which lasts only for ten days at the beginning of every school year. This course concentrates on dealing with the Communicative Approach in teaching to help those teachers to teach the Crescent English Course effectively, little is said about testing. The Inspectorate has also to convince teachers to get rid of their rigidity and begin to accept the newly added ideas on testing.

Problem-solving ability depends on real cooperation between teachers and the Inspectorate members since the desired improvement cannot be done on an individual basis. The traditional methods of test-construction should be replaced by the most recent methods and techniques in testing and the theoretical suggestions should also be moved from the theory framework to the field of practice. This change process is the key factor in the modification.

In the thesis, we have been mainly concerned with classroom testing rather than external examinations. Most chapters in this study deal with classroom tests that should be able to produce a satisfactory interpretation which directs the teaching-learning process. When we accurately test our students, we increase our own effectiveness which will affect the students' learning and provide them with an opportunity to show their abilities to produce correct forms of the target

language.

We should always keep in mind that there is a harmony between what we teach and what we test. The results we get about our students have values in the teaching-learning process. When a teacher decides to test his/her students, he/she should keep in mind certain steps to be taken before the test construction. These steps include the following:

1-Determining the purpose of the intended test.

2-Identifying and defining the intended learning outcomes.

3-Preparing the test specifications.

4-Constructing relevant test items.

5-Scoring the test

6-interpreting the test scores

The present situation in Qatar needs a modification that concentrates on adequate students' training on how to respond to certain headings. Teachers need to train their students how to deal with words such as "Distinguish-Correct-Match the sentences-Select", and the like. A large percentage of the students who fail a test lack the needed understanding of the test headings. I have noticed so many times in the past few years that many students ask about the headings during the test administration. They often ask for clarification of certain headings. When the headings are originally written in the native language, the results will certainly be better. I do

not blame the students, but teachers are to be blamed because they do not train their students on how to respond to such headings. We want the Inspectorate to continue its constant attempts to improve school examinations, but not within the confines of traditional methods. Inspectors ought to listen now and then to teachers who have the ability to suggest certain ideas to improve testing procedures.

R E F E R E N C E S

Abdur Rauf Syed Mohammad. (1988) "Culture and Reading Comprehension", Forum, Volum 26. No. 2. PP. 44-46

Adams, Gorgia Sachs. (1964) Measurement and Evaluation in Education and Guidance, New York. Holt, Rinehartart and Winston.

Adcock, C. J. (1964) Fundamentals of Psychology. Harmondsworth, Penguin Books.

Adkins, Dorothy C. (1988) Test Construction. Ohio, Charles E. Merril Publishing Co.

Al-Khuli Muhammad. (1976) English as a Foreign Language: Linguistic Background and Teaching Methods. Riyad .Riyad University Press.

Alderson, Charles J. (1981) Communicative Language Testing. Issues in Language Testing. London, The British Council.

Allen, J. P. B. and Davies Alan. (1977) Testing and Experimental Methods, Bristol, Western Printing Services Ltd.

Angel, Y. P. lee,(et al) (eds). (1985) New Directions in Language Testing. Oxford, Pergamon Press.

Baker, David. (1989) Language Testing. London, Edward Arnold.

Bagnato, Stephen J. and Neisworthy, John T. (1981) Developmental Assessment and Curricula. Maryland, Aspen Publication.

Bell, Roger T. (1981) An Introduction to Applied Linguistics. London, Batsford Academic and Education Ltd.

Berg, Harry D. (ed.), (1965) Evaluation in Social Studies. Washington, D.C., The National Council for Social Studies.

Bloom, Benjamin S, Madaus George F. and Hastings J Thomas. (1981) Evaluation to Improve Learning. New York. McGraw-Hill Book Company.

Billing, D. E. and Furniss B. S.(eds). (1973) Aims, Methods and Assessment in Advanced Science Education. London, Heyoen and Son Ltd.

Brown, Sally. (1981) A Review of Criterion-Referenced Assessment. Edinburgh, Her Majesty's Stationery Office .

Brumfit, Christopher, et al. (eds). (1981) Teaching English as a Second Language. London. Routledge and Kegan Paul

Byrd, Donald R. H. (1986) "English Language Instruction In the Gulf Arab States": A report on Programs, Curricula, Teachers and Teaching Procedures, and Students, submitted to the Gulf Arab States Educational Research Centre (GASERC) Kuwait.

Carroll, Brendan. (1982) Crescent Testing Consultancy, Doha, Interim Report.

Doyle, Kenneth, O. (1984) Evaluating Teaching. Canada. Heath and Company.

Entwistle, N. J. and Nisbet J. D. (1974) Educational Research in Action. London, University of London Press Ltd.

Finocchiaro, Mary and Bonomo, Michael. (1983) The Foreign Language Learner, New York, Regent Publication Company.

Gronlund, Norman E. (1981) Measurement and Evaluation in Teaching. 4th Edition, New York, Macmillan Publishing Company. Inc.

Gronlund, Norman E. (1985) Measurement and Evaluation in Teaching. 5th Edition, New York, Macmillan Publishing Company. Inc.

Gronlund, Norman E. (1982) Constructing Achievement Tests. Third Edition, New York, Macmillan Publishing Company, Inc.

Goodwin, William L. and Driscoll Laura A. (1980) Handbook for Measurement and Evaluation in Early Childhood. California, Jossey-Bass Publishers.

Harris, David P. (1969) Testing English as a Second Language. New Delhi. McGraw-Hill Publishing Company.

Heaten, J. B. (1985) Writing English Language Tests. Ninth Edition, Singapore, Longman Singapore Publishers Pte, Ltd.

Hopkins, Charles D. and Antes Richard L. (1985) Classroom Measurement and Evaluation. Illinois, F. E. Peacock Publishers, Inc.

Hopkins, David. (1988) A Teacher's Guide to Classroom Research. Suffolk, St. Edmundsbury Printing Press.

Hussein, Abdul Hamid. (1982) Basics in Linguistics. Jeddah. Arabia, Dar-Al Shorouq.

Karmel, Louis J. (1978) Measurement and Evaluation in the Schools, New York. Macmillan Publishing Company.

Klausmeier, Herbert J. and Goodwin. (1971) Learning and Human Abilities Singapore, Times Printers Sdn. Bhd.

Lado, Robert. (1977) Language Testing. Hong Kong, Wing Tai Cheung Printing Co. Ltd.

Lindeman, Richard, H. (1971) Educational Measurement. Bombay. R. J. Taraporevala.

Lord, Robert. (1985) New Directions in Language Testing, Oxford, Pergamon Press.

Macintosh, H. G. (1976) Assessing Attainment in the Classroom. London, Hodder and Stoughton.

Madsen, Harold S. (1983) Techniques in Testing. New York. Oxford University Press.

Madsen. (1983) "Language Testing and Current Status". Modern Language Journal, Volum 67, No. 4. PP. 431-443

Murdoch, George A. (1989) "Pragmatic Basis for Course Design". Forum. Vo. 27, No. 1. PP. 15-18

Oller, John W. (1979) Language Tests at School. London, Longman.

Palardy, J. Michael. (1973) Teaching Today. New York, Macmillan Publishing Co. Inc.

Qotbah, Mohammad. (1990) Needs Analysis and the Design of Courses in English for Academic Purposes: A Study of the Use of English Language at the University of Qatar. A thesis submitted for the degree of Doctor of Philosophy, University of Durham.

Roid, Gale H. and Haladyna, Thomas M. (1982) A Technology for Test-Item Writing. London, Academic Press.

Satterly, David. (1985) Assessment in Schools. Worcester. Billing and Sons,

Swezey, Robert W. (1981) Individual Performances Assessment and Development: An Approach to Criterion-Referenced Test Development. Virginia. Reston Publishing Company. Inc.

Sax, Gilbert. (1980) Principles of Education and Psychological Measurement and Evaluation, California, Wadsworth Publishing Co.

Thorndike, Robert L. and Hagen Elizabeth. (1969) Measurement and Evaluation in Psychology and Education. Edition. New York. John Willy and Sons Inc.

Torsten, Husen and T. Noville. (eds.) (1985) The Encyclopedia of Education, Exeter, A. Wheaton and Co. Ltd.

Tuckman, Brace W. (1973) Measuring Educational Outcomes. New York. Harcourt Brace Jovanonich, Inc.

Vallete, Rebecca M. (1077) Modern Language Testing. New York, Harcourt Brace Jovanovich, Inc.

Valdes, Joyce Merril. (ed). (1988) Culture Bound. New York. Cambridge University Press.

Weir, Cyril J. (1988) Communicative Language Testing. Exeter. University of Exeter Linguistic Studies, Vol. II

Widdowson, H. G. (1979) Exploration in Applied Linguistics. Oxford, OUP.

BIBLIOGRAPHY OF WORKS CONSULTED

Abu-Sayf, F. K. (1979) Recent Development In The Scoring Of Multiple Choice Items. Educational Review. Vol. 31, No. 3. PP. 269-278

Adams, Gorgia Sachs. (1964) Measurement and Evaluation in Education and Guidance. New York. Holt, Rinehart and Winston.

Adcock, C. J. (1964) Fundamentals of Psychology. Aylesbury, Penguin Books.

Adkins, Dorothy. (1974) Test Construction. Ohio, Charles E. Merrill Publishing Company,

Al-Khuli Muhammad. (1976) English as a Foreign Language: Linguistic Background and Teaching Methods. Riyad, Riyad University Press.

Alderson, Charles J. (1981) Issues in Language Testing. ELT Documents 111. London, The British Council.

Allen, J. P. B. and Davies Alan. (1977) Testing and Experimental Methods. Bristol. Western Printing Services Ltd.

Allen, Verginia French. (1983) Techniques in Teaching Vocabulary. New York. Oxford University Press.

Al-Misnad, SheiKha. (1985) The Development of Modern Education in The Gulf. London, Ithaca Press.

Angela, Y. P. lee, et al (eds.). (1985) New Directions in Language Testing. Pergamon Press.

Archer, John and McCarthy, Barry. (1988) Personal Biases in Student Assessment. Educational Research, Volume 30 No. 2. PP. 142-144

Bachman, Lyle F. (1990) Fundamental Considerations in Language Testing. Hong Kong, Oxford University press.

Bagnato, Stephen J. and Newsworthy, John T. (1981) Developmental Assessment and Curricula. Maryland Aspen Publication.

Balet, Sebastian. (1985) Testing Some Current Assumptions. ELT Journal. Vol. 39. No. 3. PP. 178-182

Baker, David. (1989) Language Testing. London. Edward Arnold.

Basset, Steven and Woolf, Malcolm. (1988) How Classroom Assistants Respond. British Journal Of Special Education. Vol.2, No.2. P.17

Bastidas, Jesus A. (1984) The Cloze Procedure As A Teaching Technique To Improve Reading Comprehension. Forum. Vol. 22. No. 2. PP. 20-24

Bell, Roger T. (1981) An Introduction to Applied Linguistics. London. Academic and Education Ltd.

Berg, Harry D. (ed.) (1965) Evaluation in Social Studies. Washington D.C. The National Council for Social Studies.

Berk, Ronald A. (1982) Handbook of Methods for Detecting Test Bias. London, The John Hopkins Press.

Billing, D. E. and Furniss B. S. (eds.) (1973) Aims, Methods and Assessment in Advanced Science Education, London, Heyeon and Son Ltd.

Black, Harry and Broadfoot Batricia. (1982) Keeping Track of Teaching. London. Routledge & Kegan Paul Ltd.

Bloom, Benjamin S. et al (1981) Evaluation to Improve Learning. New York. McGraw-Hill Book Company.

Borg, Walter R. & Gall Meredith D. (1983) Educational Research New York, Longman, Inc.

Borg, Walter R. (1981) Applying Educational Research. New York, Longman, Inc.

Brown, Sally. (1981) A Review of Criterion-Referenced Assessment. First Edition. Edinburgh, Her Majesty's Stationary Office.

Broughton, Geoffrey and Brumfit, Christopher. (1980) Teaching English as a Foreign Language. London. Routledge & Kegan Paul Ltd.

British Council. (1978) "Crescent Course Evaluation": A report submitted to the Ministry of Education, Doha, Qatar, PP. 1-4

British Council. (1982) "Testing the Crescent Course": A report submitted to the Ministry of Education. Doha, Qatar, PP.1-15

British Council. (1984) " Possible Ways of Improving the English Language Proficiency of Qatari Students": A report

submitted to the Ministry of Education. Doha, Qatar. PP.1-2

Bush, William Jo and Waugh, Kenneth W. (1982) Diagnosing Learning Problems. Ohio. Charles E. Merrill Publishing Company.

Byrd, Donald R. H. (1986) "English Language Instruction In the Gulf Arab States": A report on Programs, Curricula, Teachers and Teaching Procedures, and Students, submitted to the Gulf Arab States Educational Research Centre (GASERC) Kuwait.

Carroll, Brendan. (1982) Crescent Testing Consultancy, Doha, Interim Report.

Carroll, Brendan J. and Hall Patrick J. (1985) Make Your Own Language Tests. First Edition. Exeter . A. Wheaton & Co. Ltd.

Clark, Burton R. A View From 1985. (1985) Evaluation Comment. Vol. 7. No. 2. PP. 2-11

Committee of Arab Linguists. (1983) A Dictionary of Modern Linguistic Terms. Beirut. Librairie du Liban.

Council of Europe, (1988) Evaluation and Testing in the Learning and Teaching of Languages for Communication. France. Project No. 12.

Cruickshank, William M. (1980) Approaches To Learning. New York, Syracuse University Press.

David, Hopkins. (1988) A Teacher's Gide to Classroom Reseaech. Suffolk. St. Edmundsbury Press Ltd.

David, Hopkins, (1986) In-service Training Development. Worcester. and Sons Limited.

Delighton, Lee C. ed. (1985) The International Encyclopedia of Education. Exeter. A. Wheaton & Co. L.

Doyle, Kenneth O. (1984) Evaluating Teaching. Canada. Heath and Company.

Entwistle, N. J. and Nisbet, J. D. (1973) Educational Research in Action. London. The Chaucer Press Ltd.

Finocchiaro, Mary and Bonomo, Michael. (1973) The Foreign Language Learner, New York. Regents Publishing Company.

Gillham, W. E. (ed). (1973) Psychology Today.London. Hazell Waston & Viney Ltd.

Gipps, Caroline. (1988) The Debate Over Standards and The uses Of Testing. British Journal of Educational Studies Vol. 36. No. 1. PP. 21-34

Goodwin, William L. and Driscoll Laura A. (1980) Handbook for Measurement and Evaluation in Early Childhood. California. Jossey-Bass Publishers.

Gulf Arab States Educational Research Centre (GASERC) (1983) "Development of English as a Foreign Language In General Education In the Arab Gulf States": A final report presented as a seminar in Manama, Bahrain. PP. 90-97

Green, Peters (ed.). 1987) Communicative Testing. France, Council of Europe, Publication Section.

Gronlund, Norman, E. (1981) Measurement and Evaluation in Teaching. Fourth Edition, New York, Macmillan Publishing Company. Inc.

Gronlund, Norman E. (1985) Measurement and Evaluation in Teaching. 5th Edition, New York, Macmillan Publishing Company. Inc.

Gronlund, Norman, E. (1982) Constructing Achievement Tests. Third Edition, New York, Macmillan Publishing Company, Inc.

Harris, David, P. (1969) Testing English as a Second Language. New Delhi. McGraw-Hill Publishing Company.

Harry Torrance & Murphy Roger. (1988) The Changing Face of Educational Assessment. First Edition. Bristol. J. W. Arrowsmith Ltd.

Heaten, J. B. (1985) Writing English Language Tests. Singapore, Longman Singapore Publishers Pte Ltd.

Herman, Joan L. et al (1987) Evaluator's Handbook. New Delhi. Sage Publication.

Hirst, P. H. & Peters, R. S. (1970) The Logic of Education. London. Routledge & Kegan Paul.

Hopkins, Charles D. and Antes Richard L. (1985) Classroom Measurement and Evaluation. Illinois. F. Peacock Publishers, Inc.

Hopkins, David. (1986) A Teacher's Guide to Classroom Research. Suffolk. Milton Keynes.

Hughes, Arthur. (1990) Testing for Language Teachers. Glasgow. Bell & Bain Ltd.

Hussein, Abdel Hamid Ibrahim. (1982) Basics in Linguistics. Jeddah, Saudi Arabia. Dar Al Shorouq.

Hutchinson, Brian, et al. (1988) The Problem of Validity in The Qualitative Evaluation of Categorically Funded curriculum Development Projects. Educational Research, Vol. 30. No. 1. PP. 54-64

Johnson, Brian, et al. (1987) Assessing English. Suffolk. St. Edmundsbury Press.

Kahler, wolfgang. (1986) Spinning The Tale: An Approach To The Testing Of Reading Comprehension. Forum. Vol. 24. No. 4. PP. 27-28

Karmel, Louis J. (1978) Measurement and Evaluation in the Schools. New York. Macmillan Publishing Company.

Klausmeier, Herbert J. and Goodwin. (1971) Learning and Human Abilities. Singapore, Times Printers Sdn. Bhd.

Lado, Robert. (1977) Language Testing. Hong Kong, Wing Tai Cheung Printing Co. Ltd.

Laesch, Bowers & Kleeck Ann Van. (1987) The Close Test as an Alternative Measure of Language Proficiency of Children Considered for Exit from Bilingual Education Programs. Language Learning. A Journal of Applied Linguistic. Vol. 37. No. 2. PP.171-187

Lewis, D. J. (1973) Statistical Methods in EDucation. London, University of London Ltd.

Lidz, Carol Schneider. (1981) Improving Assessment of Schoolchildren. London. Jossey-Brass Publishers.

Lindeman, Richard, H. (1971) Educational Measurement.Bombay. R. J. Taraporevala.

Lord, Robert. & Low Graham.(eds.) (1985) New Directions in Language Testing, Oxford. Pergamon Press.

Murphy, Roger J. L. (1978) Sex differences in Examinations Performanc. Educational review. Vol. 30, No. 3. PP. 259-262

Macintosh, H. G. (1976) Assessing Attainment in the Classroom. London, Hodder and Stoughton.

Madsen, Harold S. (1983) Techniques in Testing. New York. Oxford University Press.

Madsen. (1983) Language Testing and Current Status. Modern Language Journal, Volum 67, No. 4. PP. 431-443

Marshal, Jon Clark. (1972) Essentials of Testing. Canada. Addison Wesley Publishing Company.

Mason, W. H. (1970) Advanced Tests in English. London. The Camelot Press Ltd.

Matthews, R. J. H. (1972) Translation as a Testing Device. English Langauge Teaching. Vol. 27. No. 1. PP. 58-63

Mersand, Joseph. (1977) The English Teacher. First Edition, London. KennKat Press.

Ministry of Education. (1990) Qatar Year Book. Doha, Qatar, Press and Publication Department.

Morrow, K. (1979) Communicative Language Testing: Revolution or Evolution. In Brumfit, C. J & Johnson, K. The Communicative Approach to Language Teaching. Oxford. OUP.

Murdoch, George A. (1989) Pragmatic Basis for Course Design. Forum. Volum 27. No. 1. PP. 15-18

Murphy Roger and Torrance Harry. (1988) The Changing Face of Educational Assessment. Bristol. J. W. Arrowsmith Ltd.

Miller, Ann. (1982) The Development of a Language Aptitude Test. Oxford. University of Oxford.

Noordhoff, Wolters. (1984) Applied Linguistics and the Learning and Teaching of a Foreign Language. Translated by Van Oirsouw R. R. London. Edward Arnold (publishers) Ltd.

Oller, John W. (1979) Language Tests at School. London, Longman.

Oller, John W. (ed.) (1983) Issues in Language Testing Research Massachusetts. Newbury House Publishers, Inc.

Palardy, J. Michael. (1975) Teaching Today. New York, Macmillan Publishing Co. Inc.

Perren, George. (1972) Testing Ability In English As A Second Language. English Langauge Teaching, Vol. 17. No. 1. P. 22.28.

Perren, George.(1966) Testing Ability in English as a Second Language. English Language Teaching. Vol. 21. No. 1. PP. 99-106

Roberts, Geoffrey R. (1973) Reading in Primary Schools. London. Unwin Brothers Limited.

Roid, Gale H. and Haladyna, Thomas M. (1982) A Technology for Test-Item Writing. London. Academic Press.

Saleemi, Anjum P. (1988) Language Testing: Some Fundamental Aspects. Forum, Vol. 16, No. 1. PP. 50-64

Satterly, David. (1985) Assessment in Schools. Worcester G. B. Billing and Sons.

Savignon, Sandra, J. (1987) What's What in Communicative Language Teaching. Forum. Vol. 15, No. 4. 16-21

Sax, Gilbert. (1980) Principles of Education and Psychological Measurement and Evaluation, California. Wadsworth Publishing Co.

Seddon, G. M. (1988) The Validity of Reliability Measures. British Educational Research Journal. Vol. 14. No. 1. PP. 89-97

Skager, Rodney, et al (eds). (1978) Curriculum Evaluation For Lifelong Education. London, Pergamon Press Ltd.

Skehan, Peter. (1988) Language Testing, Part 1. The International Abstracting Journal For Langauge Teachers and Applied Linguists. Vol. 21. No. 4. PP. 211-221

Souper, Patrick C. (1967) About To Teach. London. Routledge & Kegan Paul Ltd.

Sponcer, Frank. (1983) Testing for Teaching. First Edition. London. The Chaucer Press Ltd.

Swezey, Robert W. (1981) Individual Performances Assessment and Development: An Approach to Criterion-Referenced Test Development. Virginia. Reston Publishing Company. Inc.

Thorndike, Robert L. and Hagen Elizabeth. (1969) Measurement and Evaluation in Psychology and Education. New York. John Willy and Sons Inc.

Torsten, Husen and T. Noville. (eds) . (1985) The Encyclopedia of Education. Exeter. A. Wheaton and Co. Ltd.

Torrance, Harry. (1986) What Can Examinations Contribute To School Evaluation. Educational Review. Volum 38, No. 1. PP. 31-42

Tuckman Brace (1978) Conducting Educational Research. 2nd ed.New York. Harcourt Brace Jovanovich, Inc.

Tuckman, Bruce W. (1975) Measuring Educational Outcomes: Fundamentals of testing. New York. Harcourt Bruce Jovanovich, Inc.

Underhill, Nic. (1987) Testing Spoken Language: A Handbook of Oral Testing Techniques. First Edition. Cambridge. Cambridge University Press.

Vallete, Rebecca M. (1877) Modern Language Testing. New York, Harcourt Brace Jovanovich, Inc.

Valdes, Joyce Merril. (ed). (1988) Culture Bound. New York. Cambridge University Press.

Vernon, M. D. (1962) The Psychology of Perception. Middlesex. C. Nicholls & Company Ltd.

Walters, Jane. (1954) Techniques of Counceling. New York, McGraw-Hill Book Company.

Weir, Cyril. (1988) Communicative Language Testing. Exeter. A Wheaton & Co. Ltd.

Widdowson, H. G. (1979) Exploration in Applied Linguistics. Oxford. Oxford University Press.

William, Eddie. (1983) A Humanistic Approach to Classroom Achievement Testing. MET. Vol. 10. No. 4. PP. 31-34

Yopp, Hallie Kay. (1988) The Validity and Reliability of Phonemic Awareness Tests. Reading Research Quarterly. Vol. 23, No. 2. PP. 159-171

Zigmond, Naomi, et al. (eds.) (1983) Assessment for Instructional Planning in Special Education. New Jersey. Printice-Hall, Inc.

APPENDICES

APPENDIX (A)

PRIMARY LEVEL OBJECTIVES:

1-Listening:

1-Pupils should be able to recognise forms of greeting and introduction.

2-They should be able to understand the teacher's instruction in English.

3-They should be able to understand the teacher's classroom questions (Where is your book ? Did you do your homework ?)

4-They should be able to listen to a short simple story or a short dialogue and understand the main points (When, Where Who, how much etc.)

5-They should be able to understand the general meaning of a short passage even if a few words are unknown to them, which they should guess from the context.

2-Speaking:

1-Pupils should be able to use the basic courtesies of greetings, introductions and apology

2-They should answer the teacher's questions dealing with school and study

3-They should be able to reply to simple questions based on short written passages or pictures of people and things.

4-They should be able to ask questions requesting basic information that deals with classroom English (What is the meaning of ...? How to do this exercise...?).

5-They should be able to give simple orders and ask for things from their classmates or from their teacher .)

6-They should be able to talk about themselves and their hobbies

7-They should be willing to use English in group activities, games.

8-They should be able to perform the "Oral Performance "items at their level and to read aloud some short sentences and short paragraphs in order to practice pronunciation.

3-Reading:

1-Pupils should be able to recognise and read aloud the letters of the alphabet.

2-They should be able to read the sentences and paragraphs contained in their books

3-They should be able to recognise and read words and phrases from flash cards.

4-They should be able to manage a reader at the appropriate level and they should be able to read simple comics.

5-They should be able to use a suitable picture dictionary.

4-Writing:

1-The pupils should be able to write correctly in small and capital letters and numbers

2-The should be able to write most of the words they have learned correctly and others recognizably.

3-They should be able to write their own names and addresses, copy down those of other pupils.

4-They should be able to write a sentence in reply to a written or spoken question.

5-They should be able to compose one or two sentences of their own about familiar objects when prompted by a model.

APPENDIX (B)

PREPARATORY LEVEL OBJECTIVES:

1-Listening:

By the end of the preparatory level, the pupils should be able to do the following :

1-Listen to orders and instructions and carry out a talk (playing a game, filling in a form, translating directions etc.)

2-Listen to authentic English transmitted through the media or airport announcement or a telephone conversation, and be able to extract specific information.

3-Listen and understand short extracts from the media like the news commentaries and plays.

4-Listen to and identify the basic theme of a short conversation between two speakers, as well as their opinions and kind of relationship they have (friendly or otherwise).

5-Understand the main idea and important points of a story or a dialogue.

6-Identify the situation from the context of a conversation

between two people (a visit to the doctor, buying something from a store etc.)

2-SPEAKING:

By the end of the preparatory level, the pupils should be able to express themselves in a clear and acceptable way in different situations such as:

1-Taking part in a short social conversation concerning home, school, travel, media and local environment.

2-Copying with English language contact situations as helping a foreigner in the market or the street;giving him simple direction; giving explanations of situations related to the local environment.

3-Handling language situations in a foreign country where English is the means of communications.

4-Taking part in a telephone conversation to give an invitation, make an apology or question a certain personal or practical situation.

5-Telling a story or describing an accident or relating an incident.

6-Passing a telephone or personal message or reporting information from a written or spoken source.

3-Reading:

By the end of the preparatory level, the pupils should be able to read and understand the following:

1-A short story or expository passage in simple English with only a few unknown words, making an intelligent guess at the meaning of unknown words from their context.

2-Extract information from a letter or an article and act accordingly.

3-Interpret and follow written instructions and act accordingly

4-Understand the news, reports and advertisements.

5-Extract information from an alphabetical list(telephone directory and interpret maps, diagrams, tables and graphs.

6-Use simple dictionaries.

4-Writing:

1-Pupils should be able to write about themselves, their families and school.

2-They should be able to summarise a spoken story or write descriptions of incidents they have witnessed.

3-They should be able to write letters (e.g.to a pen-friend) asking for information or advice.

4-They should be able to fill in forms such as those for passports at the airport.

5-They should be able to write down a simple summary of information from a map or a diagram.

APPENDIX (C)

SECONDARY LEVEL OBJECTIVES:

1-Listening:

By the end of the secondary stage, pupils should be able to understand spoken English in various situations such as :

1-A conversation on a general topic between speakers of various accents.

2-A conversation on a familiar topic between two native speakers so as to be able to report it to a third person.

3-English through the audio-visual media e.g. radio, TV, news broadcasts, plays and films.

4-A non-academic talk, e.g., by a tourist guide.

5-A short talk (mini-lecture) taking notes while listening.

6-Follow a set of instructions and carry them out, for example, read a map, fill in a form or questionnaire, carry out a scientific experiment.

2-Speaking:

By the end of the secondary stage, pupils should be able to communicate clearly and specifically in English in various situations such as :

1-Taking part in a short conversation without a major breakdown

in communication.

2-Giving a short talk from notes.

3-Giving instructions involving a series of points.

4-Summarizing verbally a written or spoken message or report.

5-Narrating a story or incident .

6-Expressing interests and personal attitudes.

7-Talking on either literary or scientific topic in a media-style interview

8-Giving details of an incident (e.g., car accident , theft); describing symptoms of illness and requesting treatment.

9-Commenting on a play.

10-Acting out a sketch with other pupils, role play.

3-Reading:

By the end of the secondary stage, pupils should be able to use their reading skills to do the following:

1-Consult reference books;

2-Read and understand a straightforward passage of about 300 words containing some unfamiliar information and vocabulary.

3-A suitable original novel.

4-Personal and business letters.

5-Cables, advertisement, common abbreviations, menus, theatre and sports programmes, tourist brochures etc.

6-Written instructions on, for example, how to use electrical appliances, take medicine.

7-Work out the meaning of unknown words and phrases from a specific context.

8-Extract information from graphs. charts, and references.

4-Writing:

By the end of the secondary stage, pupils should be able to

1-Express their points of view on topics of personal or scientific interest.

2-Organise a report on information from different sources.

3-Write detailed personal and business letters using correct formats

4-Fill in forms of different sorts, such as those for passports, airport, customs clearance, entrance to university or institute.

APPENDIX (1 A)

DESCRIPTION OF THE COURSE MATERIALS:

The materials are produced by the Oxford University Press (OUP) in Beirut, Lebanon. They are divided into teacher's materials and pupil's materials.

Primary Five: (Year 1: Five periods per week).

Teacher.	Pupils.
1. Teacher's Book 1.	Pupil's Book 1.
2. Teacher's Cassette 1A, 1B.	Pupil's Cassette 1.
3. Teacher's Pack 1 (flashcards).	Pupil's Workbook 1.
4. OHP Transparencies.	Pupil's Handwriting Book 1&2

APPENDIX (1 B)

PRIMARY SIX: (year 2: Five periods per week)

Teacher.	Pupils.
1. Teacher's Book 2.	Pupil's Book 2.
2. Teacher's Cassette 2A, 2B.	Pupil'S Cassette 2.
3. Teacher's Pack 2 (flashcards)	Pupil's Workbook 2.
4. OHP Transparencies.	Pupil's Handwriting Book 1&2

APPENDIX (2)

PREPARATORY ONE: (year 3: six periods per week).

Teacher	Pupils
1. Teacher's Book 3.	Pupil's Book 3.
2. Teacher's Cassette 3A, 3B.	Pupil's Cassette 3.
3. Wallsheets.	Pupil's Workbook 3.
4. OHP Transparencies.	Pupil's Handwriting Book 1&2

APPENDIX (3)

PREPARATORY TWO: (year 4: Six periods per week).

Teacher	Pupils
1. Teacher's Book 4.	Pupil's Book 4.
2. Teacher's Cassette 4A, 4B.	Pupil's Cassette 4.
3. Wallsheets.	Pupil's Workbook 4
4. OHP Transparencies.	Oxford E.A. Dictionary.

APPENDIX (4)

PREPARATORY THREE: (year 5: Six periods per week).

Teacher	Pupils
1. Teacher's Book 5.	Pupil's Book 5.
2. Teacher's Cassette 5A, 5B.	Pupil's Cassette 5.
3. Wallsheets.	Pupil's Workbook 5
4. OHP Transparencies.	Oxford E.A. Dictionary.

APPENDIX (5)

SECONDARY ONE: (year 6: Six periods per week).

Teacher	Pupils
1. Teacher's Book 6.	Pupil's Book 6.
2. Teacher's Cassette 6A, 6B.	Pupil's Cassette 6.
3. OHP Transparencies.	Pupil's Workbook 6.
4. ----------E. A. Reader's Dictionary	

APPENDIX (6)

SECONDARY TWO: (year 7: Arts).

Boys: 8 periods per week (one period translation).

Girls: 7 periods per week (one period translation).

Teacher	Pupils
1. Teacher's Book 7.	Pupil's Book 7.
2. Teacher's Cassette 7A, B, & C.	Pupil's Cassette 7A, 7B.
3. OHP Transparencies.	Pupil's Workbook 7.
4. ----------E. A. Reader's Dictionary.	

* Extra for Science Section 1.
(Teacher's Book 1). (Pupil's Book 1). (Teacher's Cassette 1).

APPENDIX (7)

SECONDARY THREE: (year 8)

Arts: 8 periods per week. (one period translation).

Science 6 periods per week. (one period translation).

Teacher	Pupils
1. Teacher's Book 8.	Pupil's Book 8.
2. Teacher's Cassette 8A, B, & C.	Pupil's Cassette 8A, 8B.
3. OHP Transparencies.	Pupil's Workbook 8
4. ----------E. A. Reader's Dictionary.	

* Extra for Science Section 2.

(Use Scientific English 2).

(Teacher's Cassette 2). (Pupil's

APPENDIX (8)

DESCRIPTION OF COURSE CONTENTS

I shall try to summarize the course contents in the following figure.

Course contents:

Level	No. of Units	No. of Steps
1-Primary 5	04	105
2-Primary 6	04	078
3-Preparatory 1	04	100
4-Preparatory 2	08	094
5-Preparatory 3	09	073
6-Secondary 1	12	106
7-Secondary 2	12	118
8-Secondary 3	08	098

APPENDIX (9)

MARKS ALLOCATION: (Monthly exams)

1 Oral Performance. (Listening and speaking).

Level	**Marks allocated**
1-Primary	06
2-Preparatory	08
3-Secondary	10

APPENDIX (10)

2. WRITTEN EXAMS. (for all levels)

Level	**Marks allocated**
1-Primary	24
2-Preparatory	32
3-Secondary 1	32
4_Secondary 2 & 3 Arts	50
5-Secondary 2 & 3 Scientific	32

APPENDIX (11)

3. MARKS DISTRIBUTION ON WRITTEN EXAMS: (for all levels).

Level	Reading	Writing	Vocab	Grammar	Textbook	Translation
Prim.	06	06	06	06	*	*
Prep.	08	08	06	06	04	*
Seco.Sc.	08	08	04	04	04	04
Seco.Ar.	11	11	07	07	07	07

Key:

Prim.= Primary.

Prep.= Preparatory.

Seco.= Secondary.

Sc. = Scientific.

Ar. = Arts.

* = No marks are allocated.

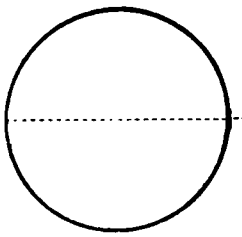
APPENDIX (12)

Ministry of Education
QATAR

Secret No.

Subject : .....

Mark



.....

No.	Marker	Reviser
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
Totalling		
Final Revision		

FINAL REMARKS -

(*In Figures*)

(*In Words*)

MINISTRY OF EDUCATION — PROMOTION EXAMINATION

PRIMARY 6 — Romilah School (girls) — Time : 2½ Hours

==========

اكتب المعلومات في البطاقة التالية :

1- FILL IN THE FOLLOWING CARD :

Name : .....

Age : .....

Weight : .....

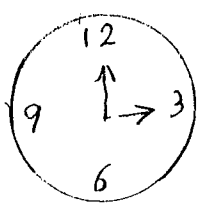
Height : .....

(2 M.)

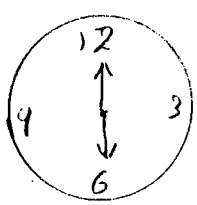
****

أكمل بوضع الزمن المناسب

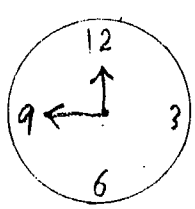
2- Cmplete the suitable time .



It is three .....



It is six .....



It is nine .....

(3 M)

==========

أكمل الجمل التالية :

3- Complete The Following Sentences.

1-We are playing in the .....

2-Yesterday I ate a .....

3- We .....to school (3 M.)

==========

اكتب عكس ما يلي :

4- Give the ppposites .

1- black .....

2- late .....

3- old .....

4- sad .....

(8 M.)

اكتب الأعداد التالية بالحروف

5- Write the following numbers in letters :

1970 .....

7459 .....

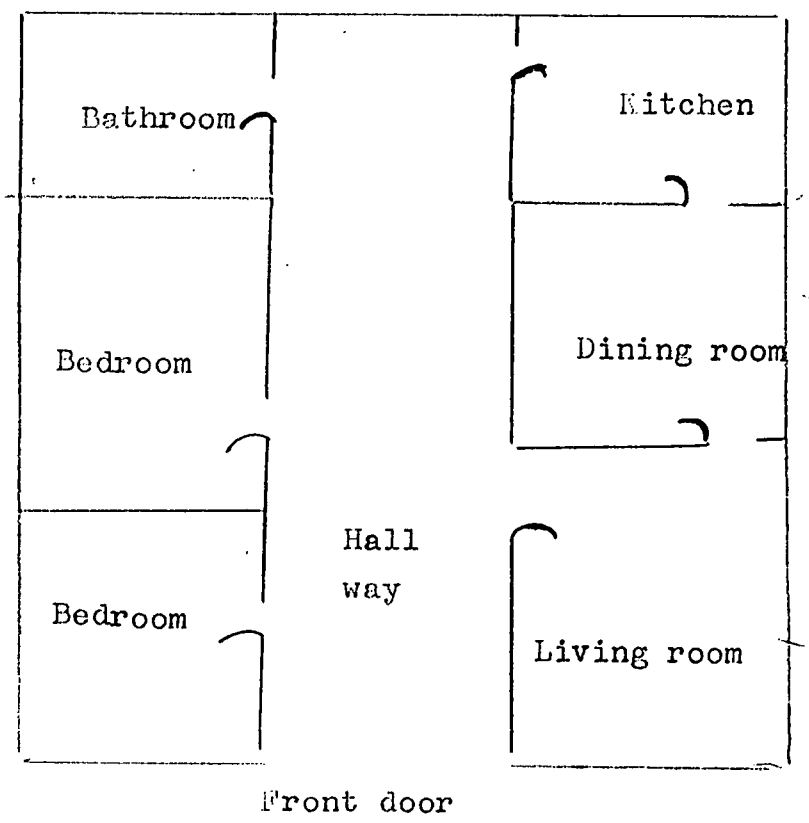
2000 ..... (3 M,)

==========

6- Describe this map of Ahmad's house

صف مخطط بيت أحمد

(2)



The house has .....

....

.... ,, ,, ...

.....

==========

قرأت في الصحف أن طفلين فقدا في مدينة كبيرة . صفهما للبوليس

7- You have read in the newspaper that two boys were lost in a big city . Describe the two boys to the police . (2)

.....

.....

.....

دعيت الى حفلة سامي وأنت -3- لا تريد الحضور . اكتب جواباً لسامي .

8- You are asked to come to Sami's party, You do not want to come write a reply to Sami .

Dear Sami ,

.....

.....

.....

.....

.....

Yours

Ahmed

(1 M.)

طبعت بمطبعة وزارة التربية والتعليم
دولة قطر ــ الدوحة

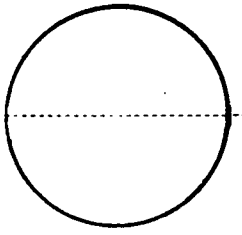
APPENDIX (13)

Ministry of Education
QATAR

Secret No.

Subject : .....

Mark



No.	Marker	Reviser
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
Totalling		
Final Revision		

FINAL REMARKS -

(*In Figures*)

(*In Words*)

STATE OF QATAR

Ministry of Education

GENERAL PREPARATORY EDUCATION CERTIFICATE AND RELIGIOUS PREPARATORY EDUCATION CERTIFICATE EXAMINATION

SECOND SESSION — 1402/1403 - 1982/1983

ENGLISH (WRITTEN) — Time: 2 Hours

١- اقرأ ما يأتي ثم اذكر الأخطار :

1. Read the following, then list the dangers:

"Accidents can be prevented. Already this year more than five hundred people have been killed or injured in accidents in their homes. Electric kitchen knives work fast, but they can cut fingers as well as food! And gas cookers may have improved life in the kitchen, but gas can cost you your life. If you think you smell a gas leak, don't light a match ..... Electricity is dangerous too, if not used correctly. For example, you should never touch switches if your hands are wet. But of course even if you are a very careful person, you could still fall downstairs and break a leg. A home can be a dangerous place!"

List the dangers which can cause accidents in the home:

1. _____

2. _____

3. _____

4. _____

٢- أ : اقرأ ما يأتي ثم أجب على الأسئلة :

2. (A) Read the following, then answer the questions:

SERGEANT EDWARDS: You're not a very good thief, are you, Harry? Imagine falling and breaking a leg while you were hard at work robbing a house! Crime is a dangerous business, isn't it?

HARRY : It wasn't the best time to have an accident, was it? I couldn't see in the dark, that's what it was.

SERGEANT EDWARDS : Knowing you, Harry, I'm surprised you didn't switch all the house lights on ..... Tell me what happened, exactly.

HARRY : Well, I saw that the family - Brown was the name, wasn't it? - were going out for the evening. I followed them to a cinema, and after I had made sure they got in, I went back to their house and broke a window. I climbed into the house. Couldn't see a thing, of course.

SERGEANT EDWARDS : What happened next?

HARRY : I was on my way upstairs to the bedroom - that's where you usually find jewellery, you know, sergeant - when I fell over the edge of a carpet. I fell downstairs and broke my leg. It was awful ..... I could hardly move. I had to phone for an ambulance. Crime doesn't pay!

SERGEANT EDWARDS : You're right there, Harry. By the way, I see you've got a cut on your hand .....?

HARRY : The broken glass, Sergeant. No, don't laugh .....

1. Why does Sergeant Edwards say "Crime is a dangerous business"?

---

2. a) How did the Brown family spend the evening?

---

b) Why did Harry follow them?

---

3. What was Harry interested in stealing?

---

4. Why didn't Harry leave the house?

---

5. How did Harry cut himself?

2.(B) What happened next? Write what you think happened after Harry telephoned for an ambulance: ٢- ب : ما الذى حدث بعد ذلك ؟
اكتب ما تعتقد أنه حدث بعد أن طلب هارى سيارة الأسعاف .

---

3. Read Sergeant Edward's interview with Harry again, then write the sergeant's report. (These words will help you - ask, reply, admit, confess) ٣- اقرأ المقابلة بين المفتش اد وارد وهارى مرة اخرى ، ثم اكتب تقرير المفتش اد وارد .

REPORT OF INTERVIEW WITH HARRY GREEN

٤- اقرأ ما يقوله المخبر الصحفي ، ثم اكتب رد مسز براون .

4. Read what the reporter says, then write Mrs. Brown's reply:

"Mrs. Brown, I'm a reporter with the Morning News. I'd like to ask you about what happened last night when Harry Green broke into your house. What did it feel like to come back from the cinema and see a police car and an ambulance outside your home? Did Harry say anything to you? How did the rest of the family feel? Oh - and Mrs. Brown - which film had you been to see?"

Now write Mrs. Brown's reply:

---

5. Help the reporter by putting these headlines in the right part of the newspaper:

٥ - ساعد الصحفي بوضع العناوين التالية في المكان المناسب من الصحيفة .

More oil found in Gulf

Doctor at Hamad Hospital Wins Prize

THIEF BREAKS LEG AT WORK!

New Programmes on Channel 37

Qatar's Team to be Announced Today

CLASSICAL MUSIC REVIEW

1. Health News: _____
2. Crime Section: _____
3. TV Page : _____
4. The Arts: _____
5. Business News : _____
6. Sports Page: _____

طبعت بمطبعة وزارة التربية والتعليم
دولة قطر ــ الدوحة

APPENDIX (14)

Ministry of Education
QATAR

Secret No.

School Year : .....

Session : .....

Year : .....

Subject : .....

Date : .....

Mark

No.	Marker	Reviser
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
Totalling		
Final Revision		

FINAL REMARKS -

(*In Figures*)

(*In Words*)

KHALIFA SECONDARY SCHOOL

ENGLISH EXAM SECONDARY ONE (PAPER TWO)

Name .....Class .....

I. Change these sentences into Indirect Speech :

1. " I believe that people have seen something " .said Fauzia .

2. "I don't believe that a ufo come from another world ".said Ali .

1.....

2. .....

II. Complete the following sentences :

1. If I were you , I . . . . . .

2. If I have money , I . . . . . .

****

III. Give the opposite of :

belief : ..... quiet : .....

true : ..... slow : .....

IV. Fill in the spaces with any suitable words :

1. .....are signs of disease .

2. UFOS produce a .....humming noise .

****

V. Translate into Arabic :

The film was really great.There were lots of fights between strange blue creatures in flying saucers and us Earth people .The film was frighening but I like that .

.....

.....

****

B) Translate into English:

يعتقد الكثيرون بأن قضاء يوم أو إثنان في الصحراء هو مضيعة للوقت. هل تصدق هذا؟

. . . . . .

. . . . . .

. . . . . .

IV. Complete this phone call between Jasim and the nurse:

Nurse: This is Hamad Hospital. Good morning.

Jasim: Good morning. . . . . .

. . . . . .

Nurse: Sorry. Dr. Abdullah is out. May I know who is speaking, please?

Jasim: . . . . . .

Nurse: Jasim Ali. OK. What can I do for you, Jasim?

Jasim: Would you please . . . . . .

. . . . . .

Nurse: OK. I'll tell the doctor that you need him immediately. Anything else?

Jasim: Yes. Tell him that . . . . . .

. . . . . .

Nurse: Your father has been vomiting all morning. Oh dear!

. . . . . .?

Jasim: 846219

Nurse: . . . . . .?

Jasim: Yes, that's all. Thank you.

V. Complete in a meaningful way:

1. I'm convinced . . . . . .

. . . . . .

2. It's very unlikely . . . . . .

. . . . . .

J.S.J.