

**People's Democratic Republic of Algeria**  
**Ministry of Higher Education and Scientific Research**  
**Mohammed Seddik Ben Yahia University -Jijel-**  
**Faculty of lettres and languages**  
**Department of English**



## **An Evaluation of Achievement Tests in Secondary School Education**

**A Dissertation Submitted in Partial Fulfillment for the Requirements  
of a Master Degree in English Didactics**

**Submitted by:**

**Naoual Mecif**

**Samira Denche**

**Board of Examiners:**

**Chairwoman: Mrs. Bouhali Selma**

**Examiner: Ms. Melit Ilham**

**Supervisor: Mrs. Lekkat Sarra**

**Supervised by:**

**Mrs. Lekkat Sarra**

**University of Jijel**

**University of Jijel**

**University of Jijel**

**Academic Year: 2018-2019**

**Dedication**

*In The Name of Allah, the Most Compassionate, the Most Merciful*

I am grateful to dedicate this work:

*To the two candles who always enlighten my life, the reason of what I become today. Thanks for your support through all my life's challenges and decisions, **dad** and **mom**.*

*To my dearest sisters; "**Wassila, Sabrina, and Farida**" who supported me all the way since the beginning of my studies.*

*To my beloved sister and her husband "**Amina**" and "**Mohammed**".*

*To my beloved brothers "**Housseem, Tarek, and Monsif**".*

*To my beautiful niece "**Rama**".*

My loyal, truthful, caring, and loving friends who supported me in every Second; **Hoda, Fatima, and Lamia**.

I cannot shut the list of the special names without naming the special friends: **Sara, Aicha, Meriem**, and all those who love me.

*To my wonderful friend and my work-partner with her, I have shared this work, university life with its lights and shadows, and my nice times; "**Samira**" and I say you are a rare treasure. I love you so much my dearest friend.*

*To all those I have forget to mention but they are always in my Heart, thank you all for your unforgettable support.*

*May God bless you all.*

*Naoual*

**Dedication*****In The Name of Allah, the Most Compassionate, the Most Merciful***

I dedicate this humble piece of work:

To the source of my success, to the one who makes my dreams real, to the beat of my heart, to the sign of challenge, the one who has drawn to me the path of success and has urged me to discover the world throughout her eyes; *to my beloved Mother.*

To the light of my eyes, to the world of compassion, to the entire beauty of this universe, to my paradise, to the holy gift I have ever; *to my dear Father.*

*To my sweetheart sisters; Nawal, Zobida, Nassira, and Ouarda.*

*To my beloved brothers; Brahim, Nabil, Madjid, Zohir, and Aziz.*

*To my wonderful nieces and nephews; Amani, kaouter, Chemseldine, Asma, Alaa, Nadjat, Djana, Meriem, Assil, Aridj, and Mohammed Takieldine.*

*To my extended family*, who have been so supportive and encouraged the fulfillment of this work.

To my lovely friend *Chahrazed* for her kindness, love, support, and care...thank you for being in my life.

To my best friends forever, *Lamia, Rima, Hoda, Sara, Farida, Habiba, Ilhem, Nabila, Meriem, and Aicha* whom I shared with unforgettable memories.

To the rose of my life and my partner in this work; *Naoual*, for her priceless words and support, and I say, *I Love you so much and I could never replay all your acts of kindness. But my heart will remain grateful to you. Thank you my pretty friend.*

To all those who believed in me and prayed for my success.

*SAMIRA*

### Acknowledgements

*"No one walks alone, and when you are walking on the journey of life ... you have to start to thank those that joined you, walked beside you, and helped you along the way"*

**David H. Hooker**

First and foremost, our utmost gratitude goes to God Almighty for His kindness in blessing us with good health, strength, and most importantly help and guidance to successfully complete this humble work.

We would like to express our extreme and hearty thanks to our supervisor Mrs. **Lekkat Sarra** for her precious guidance, perfect sense of understanding, infinite support, encouragement, and patience. We will be always in debt to her unlimited care and overwhelming kindness. **Thank you very much indeed. We totally appreciate your efforts.**

We would also like to extend our special thanks to the honorable members of the jury, Mrs. **Bouhali Selma** and **Melit Ilham** who accepted to examine and evaluate this dissertation.

Special thanks also go to our teacher Miss **Alaboudi Dalila** for her advice, help, support, and for assisting in the practical part of our study. We would not forget, of course, to express our sincere thanks and appreciation to all secondary school teachers who have kindly participated in our study, otherwise, this work would not have been possible.

We would also avail ourselves of this opportunity to extend our sincere thanks to the other teachers, from whose teaching and instruction we obtained for a life time.

**Abstract**

The present study aims at evaluating the overall usefulness of achievement tests of English in the Algerian secondary schools. In precise terms, it seeks to investigate whether achievement tests reflect the students' actual performance in relation to the context of the syllabus and their subsequent instructional objectives. For the purpose of verifying the validity of the research hypothesis, which states that achievement tests would fit the requirements of a useful test if they were well designed and developed, a descriptive research method was employed for data collection and interpretation. This investigation was carried out by the administration of a questionnaire to twenty teachers of English in different secondary schools at the district of El-Milia in Jijle. The results obtained revealed that achievement tests for EFL classes in El-Milia region do not fit the requirements of a useful test. In addition, it was found that the achievement tests are not appropriate to the CBA. At the end, the study proposes some pedagogical recommendations relevant to the design, score, and development of the achievement tests in secondary school education.

*Key words:* Evaluation, achievement test, test design, test development

**List of Abbreviations and Symbols**

ALTE: Association of Language Testers in Europe

AS: Année Secondaire (Secondary School)

BAC: Baccalaureate Examination/ Baccalaureate

CRV: Criterion Related Validity

EFL: English Foreign Language

N: Number

P: Page

TOEFL: Test of English as a Foreign Language

&: And

½: Half

¼: Quarter

%: Percentage

i.e.: That is to Say

**List of Tables**

Table 1: The Teachers' Age.....	67
Table 2: Teachers' Gender.....	68
Table 3: Teachers' Qualification.....	68
Table 4: Teachers' Teaching Experience.....	69
Table 5: The Experience of Teaching English at the Current Secondary School.....	70
Table 6: The Levels Taught by Teachers .....	71
Table 7: The Use of the Achievement Tests in Teaching.....	72
Table 8: The Teachers' Perception about the Students' Scores.....	73
Table 9: The Place of Listening and Speaking in Language Testing.....	74
Table 10: The Teachers' Reactions during the Design of the Achievement Tests.....	75
Table 11: The Place of Language Testing Design in the Educational Training.....	75
Table 12: The Place of Educational Training in Language Testing .....	76
Table 13: The Teachers' Perception toward the Need for the Special Training in Designing the Achievement Tests.....	77
Table 14: The Use of Textbook Activities in Designing the Test.....	78
Table 15: The Teachers' Use of Extra Materials .....	78

Table 16: The Procedures that the Teachers Rely on when Designing Achievement Tests.....	79
Table 17: The Teachers' Perception about Testing Procedures in the Algerian Secondary School.....	81
Tables 18: The Test Fulfillment of the Course Objective.....	82
Table 19: The Teachers' Perception toward Scoring Tests.....	83
Table 20: The Teachers' Perception toward Assessing Students' Achievement.....	84
Table 21: The Appropriateness of the Achievement Test.....	85
Table 22: The Effectiveness of Test on Students' Attainment.....	86
Table 23: The Students' Score in the Test.....	87
Table 24: The Teachers' Perception about Following the Same Scoring Procedures.....	88
Table 25: The Teachers' Perception about the Way of Assessment in the Algerian Secondary School System.....	89

**List of Figures**

**Figure 1:** The relationship between testing, assessment and teaching.....22

**Figure 2:** The test development process.....43

**List of Contents**

Dedication.....	1
Acknowledgements.....	3
Abstract.....	4
List of Abbreviations and Symbols.....	5
List of Tables.....	6
List of Figures .....	7
Table of Contents .....	8

**General Introduction**

1. Background of the Study.....	15
2. Statement of the Problem .....	16
3. Aim of the Study.....	16
4. Research Questions .....	17
5. Hypothesis.....	17
6. Research Methodology .....	17
7. Structure of the Study .....	18

**Chapter One: An Overview about Language Testing**

Introduction.....	19
1.1. Definitions.....	19
1.1.1. Definition of Tests.....	19
1.1.2. Definition of Language Tests.....	20

1.2. Related Terms.....	21
1.2.1. Test, Assessment, and Teaching.....	21
1.2.2. Test, Task, and Quiz.....	23
1.2.3. Language Testing, Language Teaching, and Language Learning.....	24
1.3. Types of Tests.....	24
1.3.1. Cloze Test.....	25
1.3.2. Diagnostic Test.....	25
1.3.3. Discrete Point Test.....	27
1.3.4. Language Aptitude Test.....	27
1.3.5. Placement Test .....	28
1.3.6. Proficiency Test.....	29
1.3.7. Progress Test.....	30
1.3.8. TOEFL Test.....	30
1.4. The Importance of Testing.....	30
1.5. Characteristics of Good Tests.....	31
1.6. Principles of Testing.....	31
1.7. The Criteria of a Useful Test.....	32
1.7.1. Validity.....	32
1.7.1.1. Face Validity .....	34
1.7.1.2. Content Validity.....	34
1.7.1.3. Criterion Related Validity .....	34

1.7.1.4. Construct Related Evidence.....	34
1.7.2. Reliability.....	34
1.7.2.1. Student Related Reliability .....	36
1.7.2.2. Rater Reliability.....	36
1.7.2.3. Test Administration Reliability.....	36
1.7.2.4. Test Reliability.....	36
1.7.3. Practicality.....	37
1.7.4. Authenticity.....	37
1.7.5. Wash back.....	38
1.7.6. Interactiveness.....	39
1.7.7. Impact.....	39
1.8. Competency Based Assessment.....	39
1.8.1. Definition of the Competency Based Assessment.....	40
1.8.2. Principles of Competency Based Assessment.....	40
1.8.2.1. Validity Assessment.....	40
1.8.2.2. Reliability Assessment.....	41
1.8.2.3. Flexibility Assessment.....	41
1.8.2.4. Fairness Assessment.....	41
1.9. Reasons for Testing.....	41
1.10. Designing a Test.....	42
1.11. Stages of Test Development.....	43

1.11.1. Planning.....	43
1.11.2. Design.....	43
1.11.3. Try-out.....	44
1.12. Payoffs and Pitfalls of Test.....	44
1.12.1. Payoffs.....	44
1.12.2. Pitfalls.....	44
Conclusion .....	44

## **Chapter Two: The Achievement Test Evaluation**

Introduction.....	46
2.1. The Historical Development of Achievement Test.....	46
2.2. Definition of Achievement Test.....	47
2.3. Types of Achievement Test.....	49
2.3.1. Teacher Made Test.....	50
2.3.2. Standardized Test.....	50
2.3.3. Performance Test.....	50
2.3.4. Final Achievement Test.....	50
2.3.5. Progress Tests.....	51
2.4. Functions of Achievement Test.....	51
2.5. Principles of Achievement Test.....	52
2.6. Planning the Achievement Test .....	52
2.7. Definition of Evaluation .....	53

2.8. Types of Evaluation .....	54
2.8.1. Formative Evaluation .....	55
2.8.2. Summative Evaluation .....	55
2.8.3. Illuminative Evaluation .....	56
2.8.4. Process Evaluation .....	57
2.8.5. Product Evaluation .....	58
2.8.6. Predictive Evaluation Vs Retrospective Evaluation.....	58
2.8.7. Impact Evaluation.....	59
2.9. Purposes of Evaluation.....	60
2.10. Participants in the Evaluation Process.....	60
2.11. The Role of Classroom Testing.....	61
2.11.1. Definition of Course Objective.....	61
2.11.2. Stimulation of Student Progress.....	61
2.11.3. Evaluation of Classroom Achievement.....	62
2.12. Test Evaluation.....	62
Conclusion.....	63

### **Chapter Three: The Field Works**

Introduction.....	64
3.1. Research Design.....	64
3.2. Participants.....	64
3.2.1. Population.....	64

3.2.2. Sampling.....	65
3.3. Data Collection Instrument.....	65
3.4. Data Collection Procedures.....	65
3.4.1. Description of the Questionnaire.....	65
3.4.2. Analysis and Interpretation of the Questionnaire .....	67
3.5. Discussion of the Results.....	90
3.6.Limitations of the Study.....	92
3.7. Recommendations and Pedagogical Implications.....	93
3.8. Suggestions for Further Research.....	94
Conclusion.....	94
<b>General Conclusion .....</b>	<b>95</b>

## **References**

## **Appendices**

## **Résumé**

## **الملخص**

## **General Introduction**

1. Background of the Study
2. Statement of the Problem
3. Aim of the Study
4. Research Questions
5. Hypothesis
6. Research Methodology
7. Structure of the Study

## EVALUATION OF ACHIEVEMENT TESTS

### General Introduction

#### Background of the Study

Language tests are regarded to be a valuable tool for providing information that is relevant to several concerns in language teaching. They can be one way of providing systematic feedback for both teachers and students. The teacher can see how well or badly the students are performing. Likewise, the students can know how much attainment and progress they are doing in learning the language.

In addition, tests can also be a good means in monitoring the learning progress or evaluating the educational system as a whole (Alderson & Buck, 1993). Shohamy (2001) supported this idea; he claimed that testing has come to play a very powerful and influential role in people's lives. This is because its results can create winners and losers, successes and failures, rejections and acceptances. (p.113) As a result, tests will have positive consequences for the stakeholders and serve as gateways or door-openers to different opportunities and position, if they are administered for the purposes for which they have been designed. However, if these instruments are used for purposes other than they have not been intended for, they can have detriment consequences on test takers serving as gate keepers, limiting their chance of success, or of joining academic or occupational positions (Alderson, Clapham & Wall, 1995; Bachman & Purpura, 2008).

It is argued that testing has evolved and has become extremely prevalent in educational systems. Consequently, all language teaching programs involve testing. Then, testing is usually done for the purpose of assessing, evaluating, and measuring the students' achievement. In other words, "The overall purpose of any form of language testing is to sample the language abilities of candidates in such a way that a realistic representation of their degree of skill in using language in non-test situations is provided" (Milanovic, 2002, p. 2).

## EVALUATION OF ACHIEVEMENT TESTS

### **Statement of the Problem**

Language testing plays an important role in assessing and evaluating the attainment of the learners in using particular language proficiency. Thus, it has been a crucial part of measurement and evaluation in the educational setting, whether its purpose is teaching or learning. Usually educational ministry or school systems identify what is expected of students when they complete a course. Therefore, the teachers in schools often design valid and reliable tests to measure the degree to which students have learned. In addition to that, the teacher as a guide supports the students to accomplish the intended outcomes and provide the learners with the appropriate feedback about how well they have learned.

Not all students achieve the same results on the test. They may succeed or fail in accomplishing the test. This may be due to the lack of ability or efforts of studying for the test. Also, it can be due to the quality of the test itself. For example, some students might poorly perform because of the misunderstanding of the tests' instructions. In other words, the tests' questions have been poorly worded or designed. That is why an understanding of how test developers come to design and develop a language test is really crucial.

### **The Aim of the Study**

The major aim of the present study is to investigate the overall usefulness of achievement tests in secondary schools. In other words, this study aims to provide an objective evaluation of the design and the development of English achievement tests in order to judge the degree of attainment and progress that students are making in accordance with the contents of the syllabus and their subsequent instructional objectives.

## EVALUATION OF ACHIEVEMENT TESTS

### **The Research Questions**

To fulfill the aim of the study, the main research question which will be explored and answered is the following:

- ✓ What are the main criteria that secondary school teachers follow to design and develop achievement tests?

In order to answer this main question, other sub-questions are needed to be answered. They are:

- ✓ What are the language skills covered by achievement tests?
- ✓ What is the degree of usefulness of these tests?
- ✓ Does the achievement test appropriate to the Competency Based Approach (CBA)?
- ✓ Do these tests take into consideration the cognitive abilities of students?
- ✓ To what extent do these tests reach the final objectives set out by teachers?

### **Hypothesis**

In the light of these research questions, the piece of research in hand hypothesizes that:

If achievement tests are well designed and developed, then the latter would fit the requirements of a useful test.

### **Research Methods and Tools**

In conducting the present study, a descriptive method which is based mainly on quantitative approach is used. To achieve the aim of the study, the data gathering tool is a questionnaire. The questionnaire, which is considered as the most appropriate means in order to figure out the main criteria for the design and the development of achievement tests, is directed to a sample of twenty secondary school teachers of English in El-Milia district in Jijel.

## EVALUATION OF ACHIEVEMENT TESTS

### **The Structure of the Study**

This research aims at providing an overall evaluation of the achievement tests for the secondary school education. In order to deal with this issue, the first two chapters constitute the theoretical part while the third chapter is the practical one.

The first chapter provides an overview about language testing including its definition, types, and the criteria of useful tests. Then, it sheds light on the competency based assessment as well as its principles. In addition, this chapter indicates the stages of designing and developing the tests.

Then, the second chapter is entitled "The Achievement Test Evaluation". It introduces achievement test; its historical development, definition, types, and principles. In addition, it deals with the evaluation including its definition, types, and purposes. The chapter ends with presenting the role of classroom testing and test evaluation as well.

The third chapter is the practical part of this research work. It deals with the analysis obtained from the teachers' questionnaire and discusses the main findings and the limitations of the study. The chapter ends by providing some pedagogical recommendations and suggestions concerning language testing.

## **Chapter One: An Overview about Language Testing**

### Introduction

#### 1.1. Definitions

##### 1.1.1. Definition of Tests

##### 1.1.2. Definition of Language Tests

#### 1.2. Related Terms

##### 1.2.1. Test, Assessment, and Teaching

##### 1.2.2. Test, Task, and Quiz

##### 1.2.3. Language Testing, Language Teaching, and Language Learning

#### 1.3. Types of Tests

##### 1.3.1. Cloze Test

##### 1.3.2. Diagnostic Test

##### 1.3.3. Discrete Point Test

##### 1.3.4. Language Aptitude Test

##### 1.3.5. Placement Test

##### 1.3.6. Proficiency Test

##### 1.3.7. Progress Test

##### 1.3.8. TOEFL Test

#### 1.4. The Importance of Testing

#### 1.5. Characteristics of Good Tests

#### 1.6. Principles of Testing

## 1.7. The Criteria of a Useful Test

### 1.7.1. Validity

#### 1.7.1.1. Face Validity

#### 1.7.1.2. Content Validity

#### 1.7.1.3. Criterion Related Validity

#### 1.7.1.4. Construct Related Evidence

### 1.7.2. Reliability

#### 1.7.2.1. Student Related Reliability

#### 1.7.2.2. Rater Reliability

#### 1.7.2.3. Test Administration Reliability

#### 1.7.2.4. Test Reliability

### 1.7.3. Practicality

### 1.7.4. Authenticity

### 1.7.5. Wash back

### 1.7.6. Interactiveness

### 1.7.7. Impact

## 1.8. Competency Based Assessment

### 1.8.1. Definition of the Competency Based Assessment

### 1.8.2. Principles of Competency Based Assessment

#### 1.8.2.1. Validity Assessment

#### 1.8.2.2. Reliability Assessment

1.8.2.3. Flexibility Assessment

1.8.2.4. Fairness Assessment

1.9. Reasons for Testing

1.10. Designing a Test

1.11. Stages of Test Development

1.11.1. Planning

1.11.2. Design

1.11.3. Try-out

1.12. Payoffs and Pitfalls of Test

1.12.1. Payoffs

1.12.2. Pitfalls

Conclusion

## **Introduction**

It is largely argued that language testing as a structured activity and a pedagogical instrument is relatively new notion that has entered the field of didactics and education. Thus, language testing is the main concern of the first chapter in our research. In this chapter, we start with presenting an overview about language testing. Then, we strive to provide definitions to the main concepts underlying testing. The first chapter also highlights the relationship between tests, assessment, and teaching, and between language testing and language teaching. In addition to that, we deal with the test types and the main criteria of testing. In addition, we refer to the competency based assessment definition and its principles. Furthermore, we introduce some reasons, steps for designing, and developing the tests. Besides, the advantages and disadvantages of tests were presented in this chapter. The achievement test evaluation will be our concern in the second chapter.

### **1.1. Definitions**

#### **1.1.1. Definition of Tests**

A test can be defined as an instrument or a procedure that is "designed to elicit certain behavior from which one can make inferences about certain characteristics of an individual" (Carroll, 1968, p. 46). Carroll maintains that the main concern of this instrument "is always to render information to aid in making intelligent decisions about possible courses of action" (p.314). However, Richards and Schmidt (2002) claimed that a test is "any procedure for measuring ability, knowledge, and performance" (p.191). In addition to that, Brown (2004) noticed that a test is "a method of measuring a person's ability or knowledge in a given domain" (p. 3). Tests yield scores that mirror attributes or characteristics of individuals (Allen, 2005). Brown's definition is considered more comprehensive because it covers all the main components of the test. Firstly, a test is a method consisting of a set of techniques, procedures, and test items that requires performance on the part of test-takers. Secondly, a test aims to measure the testee's performance. In other words,

who are the testees and what is, for example, their linguistic background knowledge. Furthermore, it measures an individual's ability or knowledge i.e. competence and know how. Lastly, a test measures a given domain or area. In the case of proficiency tests, the area is the language proficiency; for instance, communicative competence. (Brown, 2004, pp.3-4)

In brief, a test refers to the method of collecting information. Thus, it is a task or a set of tasks that use to collect and elicit observable behaviors from the test taker. In addition, tests yield the scores, and the results that represent the characteristics of the individuals. Consequently, tests are a form of measurement.

### **1.1.2. Definition of Language Tests**

Language tests refer to the formal instrument of assessment. That is, they are mainly used to assess and measure the learner's responses, language proficiency, language competence, and language abilities.

Spolsky (1989) defined language test as

"Language tests involve measuring a subject's knowledge of, and proficiency in, the of language. A theory of communicative competence is a theory of the nature of such knowledge and proficiency. One cannot develop sound language tests without a method of defining what it means to know a language, for until you decide what you are measuring, you cannot claim to have measured it ". (p. 140)

That is to say, language tests are practices used to evaluate the proficiency of students or individuals in a particular language. Therefore, they made to assess the knowledge effectively.

Bachman and Palmer (1996) defined the language test as "a tool for measuring language performance in learners" (p.10). Besides, McNamara (2000) suggested that language tests are tools that have a significant role in the educational world. As a result, language tests are used to motivate learners to study. Moreover, the tests' scores and results are used to measure the students' level, language proficiency, and determine the effectiveness of the courses. In addition, McNamara argued, "language testing is an ethical testing practice" (p.72). In other words, teachers have the responsibility on the effects of tests on the students' abilities and if the results of tests are wanted or unwanted.

To give a more specific definition to the language test, we would like to refer to Hedge (2000) definition. According to him, a test is "the specific procedures that teachers and examiners employ to try to measure ability in the language, using what learners show they know ... these tests are usually of the traditional pencil and paper variety" (p. 378).

## **1.2. Related Terms**

### **1.2.1. Test, Assessment, and Teaching**

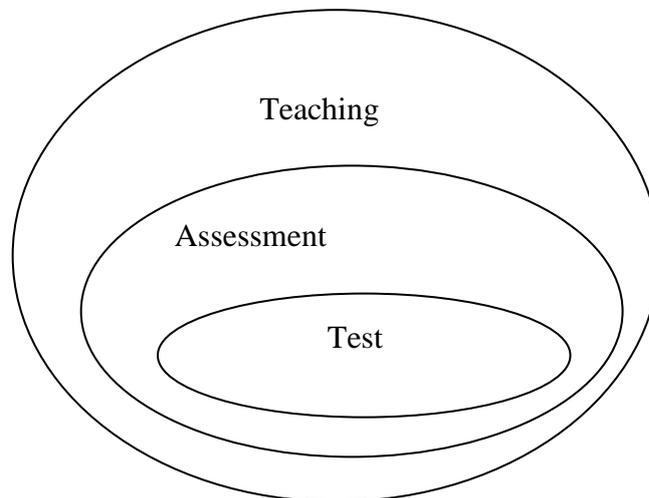
Teachers use various procedures in the classroom in order to measure their students' attainment. Among them, there are testing and assessment that take place in the process of teaching. For a nonprofessional, all these concepts may have the same use. Moreover, for some of them, they can be used interchangeably. However, for someone whose interest is educational practice or research, these particular concepts bear quite different senses. Thus, it is important to classify these terms and explain the differences between them. (Brown, 2004)

First, the terms *assessment* and *test* are used interchangeably. Brown (2004) stated that tests are administrative procedures that occur at a known time in the curriculum. Thus, the tests allow the teachers to measure and evaluate the learner's performance and responses when they master all their

faculties, capacities, and abilities. On the contrary, assessment is an ongoing process that encompasses a much wider domain. In other words, it is a process of assessing, and checking the learner's progress i.e. whenever a student responds to a question offers a comment, or tries out a new word or structure, the teacher subconsciously makes an assessment of the student's performance. Morespecifically, written workes, reading activities, and other types of language skills activities require the teacher to assess and judge the students performance. Tests are a subset of assessment i.e. the teacher can make the test as on one of the various procedures to assess the students.

The concept of teaching, on the other hand, refers to the process that includes different methods used by the teachers. In addition, it sets up the opportunities for the learners to listen, think, and takerisks, setgoals, and process feedback from the teacher.

Brown (2004) designed a diagram showing the relationship between testing, assessment, and teaching.



**Figure 1: The Relationship between Testing, Assessment, and Teaching.**

Brown's diagram shows that there is an overlap relationship between test, assessment, and teaching. That is to say, inthe process of teaching, the teacher provides instructions and activities

and makes assessment in order to evaluate the student's performance i.e., what they learn and they receive, acquire, and produce in the classroom. In addition, assessment is broader than test that is considered as a part of it. Those concepts are related to each other.

### 1.2.2. Test, Task, and Quiz

Test, quiz, and task are the main concepts of measurement in the educational world. Although these terms look similar, they are very different. They differ from one to another. First, the term *test* means a well-prepared administrative procedure that occurs at an identified time in the curriculum. Thus, it is an assessment intended to measure task-takers knowledge, skills, aptitudes, physical fitness, or classification in many other topics. In addition, the test can be occurred in written as well as spoken form. Moreover, the tests take place either as a part of the lesson or several lessons. Consequently, it takes a long period for the students to accomplish it. (Test definition in English, n.d).

Then, the concept *quiz* is defined by Oxford Learner's Pocket Dictionary as "a game in which people are asked questions to test their knowledge" (p.361). That is to say, it is a form of game in which the players or the students attempt to answer questions correctly in order to test their knowledge about a specific subject. Unlike tests, a quiz contains simple questions on the students' previous lessons. Moreover, it takes place depending on the purpose of the assessment. As a result of that, the quiz can occur daily, weekly, or monthly. Therefore, it can be used to check the understanding of students on the previous learnt knowledge.

The last term is *task*; the most famous definition of the task is given by the Oxford Advanced Learner's English Dictionary for Current English as "a piece of work that somebody has to do especially a hard or unpleasant one". That is to say, the task is an activity that focuses on the outcomes. Thus, it requires the students to use their own linguistic resources in the process to reach the outcomes.

According to Larsen-Freeman and Long (1991) "unlike the tests which are devised to measure what the learner knows and does not know of the target language, a task is devised to reveal what the learner knows " (p.41).

### **1.2.3. Language Testing, Language Teaching, and Language Learning**

Language testing is a vital component of any instructional language programme throughout the world. It almost never takes place in isolation. It is carried out for a particular purpose and in a specific content. It has a strong relationship to language teaching and learning. Upshur (1971) (as cited in Bachman, 1990, p.2) stated that there is an intrinsic reciprocal relationship between language testing and language teaching and learning. That is, language testing is served by language teaching. Consequently, language tests can be valuable sources of information about the effectiveness of learning and teaching. In addition, language testing is used as a source of information to measure different approaches to language teaching. Therefore, the language teachers regularly use tests to help diagnose students' strengths and weaknesses, to assess students' progress, and to assist in evaluating students' achievement. Besides, language testing is used as a source of feedback on learning and teaching. Thus, language tests can provide a useful input into the process of language teaching.

### **1.3. Types of Tests**

Language tests can be divided into different types depending on many criteria. Thus, tests can be classified on the basis of its purposes as well as the information they provide. Among them, there are cloze, diagnostic, discrete point, language aptitude, placement, proficiency, progress, and TOEFL test.

### 1.3.1. Cloze Test

A cloze test is a type of language testing. It refers to a cloze procedure that consists of a set of techniques for measuring reading comprehension in addition to the language proficiency. In this type of tests, words are deleted from reading passage at regular intervals, leaving blanks. Although cloze tests are associated with reading comprehension, they can be used with spoken texts as well. For instance, the transcription of spoken language (i.e. conversation) putting it in a blank. For example, every fifth word may be removed. The reader must then read the passage and try to guess the missing words. A cloze passage looks like this:

A passage used in cloze test is a ..... of written material in ..... words have been regularly..... The learners must then ..... to reconstruct the passage ..... filling the missing ..... (Richards & Schmidt, 2002, p.78).

Here, the test-taker or the reader has to guess the following missing words: a, passage, which, removed, try, by, and words.

### 1.3.2. Diagnostic Test

Another type of tests is the diagnostic test. According to Harmer (2001), diagnostic tests used to expose learner difficulties, gaps in their knowledge, and skill deficiencies during a course. Thus, when we know what the problems are, we can do something about them. (p.321) Moreover, the diagnostic tests are concerned with the achievement or lack there of. They can be used to measure the success of some small extend of teaching and to suggest where remedial work will be required.

In addition to that, Brown (2004) stated that, the diagnostic test is designed to diagnose specified aspect of a language. In other words, it seeks to identify the difficult features of a language in which learners need further help. A test in pronunciation, for example, might diagnose

the phonological features of English that are difficult for learners and should therefore become part of a curriculum. Writing diagnostic would elicit a writing sample from students that would allow the teacher to identify those rhetorical and linguistic features on which the course needed to focus on special attention. (pp. 46-47)

One of the main advantages of this test is to offer a useful pedagogical solution for mixedability classes. In this very specific context, Broughton, Barnett, and Greenwood (1980) contended that "there will certainly be a large block in the middle of the ability range that can be separated of as a group for some parts of the lesson, or for some lessons, and will form a more homogenous teaching group. If this strategy is adopted, the poor ones and the better ones must receive their due time and attention". (p.189)

One of the wellknown diagnostic tests in English is Prator's (1972) Diagnostic Passage. It consists of a short written passage that the learner reads orally, the teacher then examines a tape recording of that reading against a very detailed checklist of pronunciation errors. Diagnostic language tests have three folds objectives:

1. To provide learners with a way to start learning with their own personal learningprogramme or what would be called in the literature of testing learning paths.
2. To provide learners with a way to test their knowledge of a language.
3. To provide learners with better information about their strengths and weaknesses.

From the above, we can say the diagnostic test, as its name denotes, is designed to diagnose some particular linguistic aspects. In addition, it seeks to identify the areas in which the learners need further help. Diagnostic tests may be general. For example, if a student needs a particular help with one of the language skills. Alternatively, it may be more specific, such as seeking to determine the weaknesses, strengths, or difficulties in a student's use of past simple or one of the grammar

rules. Consequently, this type of tests is to diagnose aspects of language that the students need to develop or that a course should include.

### **1.3.3. Discrete Point Test**

Discrete point test is also known as discrete-item test. Hughes (2003) claimed that a discrete point test refers to the testing of one element at a time; item-by-item. This might involve for example, a series of items each testing a particular grammatical structure.

According to Richards and Schmidt (2002), discrete point tests refer to a type that measures knowledge of individual language items, such as a grammar test with different section on tenses, adverbs, and prepositions. They are based on the theory that language consists of different parts (eg: grammar, pronunciation, and vocabulary), and different skills (eg: listening, speaking, reading, and writing) and these are made up of elements that can be tested separately. Test consisting of multiple-choice questions are usually regarded as discrete point tests. (p.163)

Harmer noticed the following distinction between discrete-point testing and integrative testing, "Whereas discrete point testing only tests one thing at a time such as asking students to choose the correct tense of a verb, integrative test items expect students to use a variety of language at any one given time as they will have to do when writing a comparison or doing a conversational test" (Harmer, 2001, p.323). In the same line of thought, Broughton et al. noticed that, "since language is seen as a number of systems, there will be items to test knowledge of both the production and reception of the sound segment system, of the stress system, the intonation system, and morphemic system, the grammatical system, the lexical system ,and so on" (pp. 149-150).

### **1.3.4. Language Aptitude Test**

Another type of tests is known as language aptitude test. Brown (2004) stated that language aptitude tests are designed to measure capacity or general ability to learn foreign language and

ultimate success in that undertaking. Language Aptitude tests are designed to apply to the classroom learning of any language. (p. 43) That is to say, language aptitude type of test tends to measure a learner's performance in a language. Similar to that, Richards and Schmidt (2002) claimed that, "language aptitude tests refer to the test that measures a person's aptitude for second language or foreign language learning and that can be used to identify those learners who are most likely to success". They added, "a language aptitude test usually consists of several different tests that measure such abilities as;

*Sound-coding ability*, i.e. the ability to identify and remember new sounds in a new language.

*Grammar-coding ability*, i.e. the ability to identify the grammatical functions of different parts of sentences.

*Inductive-learning ability*, i.e. the ability to work out meanings without explanation in the new language.

*Memorization*, i.e. the ability to remember and to recall words, patterns, rules in the new language". (p.285)

"Two well-known standardized language aptitude tests have been used in the United States, the Modern Language Aptitude Test (Carroll and Sapon, 1958) and the Primsleur Language Aptitude Battery (Primsleur, 1966). Both of these are English tests and require students to perform such tasks as learning numbers, listening, detecting spelling clues and grammatical patterns, and memorizing". (Brown, 2004, p.43)

### **1.3.5. Placement Test**

Placement tests assess the student's level of language ability. They can be placed in an appropriate level in a course or class. Richards and Schmidt (2002) claimed that the placement test

is a test that is designed to place test-takers at an appropriate level in a programme or a course. In fact, the term *placement test* does not refer to what a test contains or how it is constructed, but to the purpose for which it is used. For instance, dictation, an interview, and a grammar test (p.404). Furthermore, Brown (2004) stated that the placement test seeks to place a student into a particular level or section of a language curriculum or school. It often includes a sampling of the material to be covered in different courses. In addition, this type of test is administered at the beginning of a new language course or at a new phase of language learning. In other words, it is intended to provide information that will help to place students at the stage of the teaching programme most appropriate to their abilities. Typically, it is used to assign students to classes at different levels. (p.45)

### **1.3.6. Proficiency Test**

One of the most significant types of test is the proficiency test. This kind of language testing is devised to measure the language ability of students at varying levels. Thus, it measures the learner's general level of language mastery. (Richards & Schmidt, 2002, p.425)

Brown (2004) claimed that the proficiency test is used to measure the students' general ability in a language. It is not limited to anyone course, curriculum, or single skill in the language; rather, it tests the overall ability. The proficiency tests may consist of standardized multiple-choice items on grammar, vocabulary, reading comprehension, and aural comprehension. They are summative and norm-referenced. In addition, Harmer (2007) claimed, "the proficiency tests have a profound backwash effect" (p .380). Moreover, McNamara (2000) stated the proficiency tests look to the further situation of language use without necessarily any reference to the previous process of the teaching. (p.7)

From the above definitions, the proficiency test administered to test a student's general ability with the language i.e. the linguistic knowledge, abilities, or skills without reference to any specific course.

### **1.3.7. Progress Test**

Progress test, as its name denotes, measures the progress that the students are making toward a defined course or program goals. This idea was supported by Richards and Schmidt (2002) who claimed that the progress tests help the teachers to judge the degree of success of his/ her teaching and to identify the weakness of the learners. In other words, the progress tests are administered at different stages throughout a course. It is used to measure the learner's progress as well as to see what they have learned. (p.427)

### **1.3.8. TOEFL Test**

TOEFL is an acronym for the Test of English as a Foreign Language.

According to Richards and Schimdts (2002) it refers to a standardized test of English proficiency administered by the educational testing service, and it is widely used to measure the English language proficiency of foreign language students wishing to enter the American universities. In addition to that, TOEFL test is a large scale of language assessment. (pp.554-555)

## **1.4. The Importance of Testing**

Keith (2001) claimed that testing is important for the teachers as well as the learners (testees). He stated that tests are helpful for the teachers because they enable them to evaluate the students' achievement and performance. In addition, tests give the teachers the opportunity to know and report their students' progress. Moreover, they allow the teachers to determine the success of

their job. Consequently, tests provide feedback for the teachers as to assess the effectiveness of the teaching methods.

On the other hand, tests are important for the learners, since they help motivate them to learn more. Therefore, learners need to know how well they are doing and learning. Moreover, the tests can help them to determine whether they understand the courses or not.

### **1.5. Characteristics of Good Tests**

Harmer (2007) summarized the main features of a good test.

A good test should:

- Be valid, reliable, and has a washback effect i.e.it should show that the tests' aims and purposes will be achieved.
- Have a well-designed purpose or intent.
- Have a foundation based on a set of written goals and objectives.
- Have a powerful effect on motivating the students to learn.
- Be a test that does the job it designed to do and which convince the people taking and making it that they work.
- Have a positive effect on both the students and teachers.

### **1.6. Principles of Testing**

Many teachers still hold a specific vision about testing. Some of them regard it as one of the most controversial areas of the teaching and learning process. It is undeniably an in-class activity that is necessary as a form of completion of the teaching input and the learning output. Basically, if properly prepared and adequately implemented, testing undoubtedly turns to be an objective pedagogical tool serving as an activity to check the effectiveness of the whole language teaching/

learning process. Test scores provide a valuable measure of how well the curriculum is being learnt and helps indicate how well students do at the main exit points of the school system, for example the baccalaureate exam.

According to Korsal (2006), to fulfill faithfully the functions that are assigned to testing, teachers should turn their attention towards the following basic principles of testing:

1. To assess learners' performance in the target language, the teacher should not give a task that the learner cannot perform. The task should be *authentic*, *realistic*, and *appropriate* to their linguistic level.

2. Even when assessing the learners' performance, at any level, the learners should be given clear instructions well. They should know what they are expected to do in a given task. The ideas, feelings, and emotions that the learners want to express cannot be limited to their insufficient linguistic input.

3. Teachers should test the outcomes or the products of what they have taught their learners, not what their colleagues know.

4. Teachers should use techniques in the teaching process as a test to have a positive washback affect of testing on language learning and teaching.

5. Teachers should test learners' writing skills by having them write and their speaking by having them speak.

6. We teach people and we evaluate language ability but we do not evaluate people.

(As cited in Benmostefa, 2014, pp.74-75)

## 1.7. The Criteria of a Useful Test

In order to judge the effectiveness of any test, it is sensible to lay down criteria against which test can be measured, as follow:

### 1.7.1. Validity

Validity is viewed as the most important criterion for any test. Thus, a good test is valid. Heaton (1988, p.169) (as cited in Kieth, 2001, p.301) defined validity as "the extent to which a test measures what is supposed to measure and nothingelse". Besides, (Gronlund, 1998,p.226) stated that " validity is the extent to which inferences made from assessment results are appropriate, meaningful, and helpful in terms of the purpose of assessment " (as cited in Brown, 2004, p.22). That is to say, the validity of tests occurs if the test tests what is supposed to measure or test.

In addition to that, Henning (1987) suggested, "validity refers to the appropriateness of a given test or one of its component parts as a measure of what purported to measure". Then, he added, "a test is said to be valid to the extent that it measures what is supposed to measure" (p. 89). On the other hand, Richards and Schmidt (2002) gave another definition to the validity as "the degree to which a test measures what is supposed to measure, or can be used successfully for the purposes for which it is intended" (p.575)

From the above definitions, although different researchers and scholars defined, the concept 'validity' in a slightly different way, but most of them argue that the test can be valid if it measures what is supposed to measure. Thus, validity means the degree to which a test accurately measures what is claimed to measure.

Harrison (1983) (as cited in Coomber, Davidson, O'Sullivan, and Stoynoff, 2012) classified validity into four kinds:

### **1.7.1.1. Face Validity**

According to (Gronlund, 1988, p.210), "face validity is the extent to which students view the assessment as fair, relevant, and useful for improving learning" (as cited in Brown, 2004, p. 26). On the other hand, Harrison (1983) defined this type of validity as "the physical appearance of the test ". That is to say, it includes the tests print, layout, and legibility. Thus, face validity refers to the extent to which the test strikes the test takers as well as the test designers, as relevant to the component that is trying to assess. Moreover, it is considered as one of the least important and the most subjective of the validity family. (As cited in Coomber, et al., 2012, p.31)

### **1.7.1.2. Content Validity**

Another type of validity is called content validity. This type of validity, as its name denotes, deals with the adequacy of the test in terms of its content. Fulcher and Davidson (2007) defined content validity as "any attempt to show that the content of the test is representative sample from the domain that is to be tested" (p.6). In other words, content validity refers to the correspondance between the test content and the content of the material to be tested.

### **1.7.1.3. Criterion Related Validity**

Criterion related validity (CRV) concerned with the degree to which the scores on newly designed test are related to those the criterion tests. Brown (2004) claimed that CRV is also known as criterion related evidence. He added that CRV refers to the extent to which the criterion of the test has actually been reached. Brown also distinguished between two types of CRV. If the developed test is administered together with criterion test concurrently that is known as *concurrent validity*. However, when the two tests are given within a time interval the relation between the two scores is called *predictive validity*. (pp. 24-25)

#### **1.7.1.4. Construct Related Evidence**

It is the most important type of validity. In the sense that, a test will not be a test if its construct validity is not established, or if it is compromised. This type of validity is concerned with psychological reality of the test. In other words, it asks the question of what it means to know language and what the nature of that knowledge is. (As cited in Coomber, etal. 2012, p.32). In short, construct validity focused on the degree in which the test items reflect the construct on which a test is based.

#### **1.7.2. Reliability**

The second criterion of tests is the reliability. (Harrison, 1983, p.10) stated, "reliability of a test is its consistency" (as cited in Kieth, 2001, p.303). Another definition of reliability was given by Bachman and Palmer (1996) who claimed, "reliability is a consistency of a measurement. A reliable test score will be consistent across different characteristics of the testing situation. Thus, reliability can be considered to be a function of the consistency or scores from one set of tests to another" (p.20). Besides, Brown (2004) argued with Bachman and Palmer, he stated that a reliable test is consistent and dependable. Consequently, the test can be reliable if the consistency of that test is similar to different students, for different times, in different places. (p.20). In addition, Richards and Schmidt (2002) stated that reliability is a measure of the degree to which a test gives consistent results. A test is said to be reliable if it gives the same results when it is given on different occasions or when it is used by different people. (p. 454)

There are different types of reliability including, student related reliability, rater reliability, test administration reliability, and test reliability.

### **1.7.2.1. Student Related Reliability**

(Mousavi, 2002, p.804) stated that the physical or the psychological factors such as fatigue and anxiety caused the most common learner related issue in reliability. (As cited in Brown, 2004, p. 21)

### **1.7.2.2. Rater Reliability**

Brown (1991) claimed that rater reliability has two types. The first type is called 'inter-rater reliability'. It happened due to the lack of attention to scoring criteria, lack of experience, and inattention. Inter rater reliability also occurs when unfair scoring criteria of two or more scorers applying the same test. On the other hand, 'intra-rater reliability' is a common occurrence for classroom teachers because of unclear scoring criteria, fatigue, bias toward particular "good " and "bad " students or simple carelessness. (As cited in Brown, 2004, p. 21)

### **1.7.2.3. Test Administration Reliability**

Another type of reliability is concerned with the administration of the test. Brown (2004) claimed that:

"there are different factors affect the unreliability of the test including photocopying variations, the amount of light in different parts of the room, variations in temperature, and even the condition of desks and chairs". (p.21)

### **1.7.2.4. Test Reliability**

Brown(2004) summarized the test reliability as follows:

The unreliability of the test occurs if the test is too long for that students may become fatigue and tired consequently they answer in a wrong way. In addition to that, the time consuming

which differentiate against good students' performance. Moreover, bad written test can make the test unreliable. (p.22)

### **1.7.3. Practicality**

Another criterion of an effective test is practicality. It is defined by Bachman and Palmer (1996) as "the relationship between the resources that will be required in the design, development, use of the test, and resources that will be available for these activities ". They added, "it is a matter of the extent to which the demands of the particular test specifications can be met within the limits of existing resources" (p.36). While, Brown (2004) stated that the practicality of a test means it is not excessively expensive. Moreover, the test is practical when it is relatively easy to administer, and has an evaluation procedure that is specific and time efficient. (p.19)

In short, the practicality of test refers to the amount of time and effort in testing. As a result, a test is practical when it is easy to design, administer, and score.

### **1.7.4. Authenticity**

The fourth important criterion of any test is the authenticity. It is considered as a major one since it describes the relationship between the test and the real world. Bachman and Palmer (1996) defined it as "the degree of correspondance of the characteristics of a given language tests task to the features of a target language task" (p.23). In addition, Brown (2004) supported this idea. He suggested the main points in which authenticity presented in a test:

- The language in the test is as natural as possible.
- Items are contextualized rather than isolated.
- Topics are meaningful (relevant, interesting) for the learner.
- Some thematic organization to items is provided, such as through a story line or episode.

- Tasks represent, or closely approximate real world tasks.

(p.28)

### 1.7.5. Wash back

Wash back is another criterion of tests. (Hughes, 2003, p.1) defined it as “the effect of testing on teaching and learning” (as cited in Brown, 2004, p.28). That is to say, wash back or backwash affects the process of teaching and learning.

More specifically, Richards and Schmidt (2002) stated that the wash back refers to the positive or the negative impact of a test on classroom teaching and learning (p. 586). In other words, the wash back affects the tests either positively or negatively. The former occurs for example if the tests encourage learners to learn more and more. The latter takes place for instance if the tests have unexpected or harmful results.

Shohamy, Donitsa-Schmidt, and Ferman (1996) defined wash back very simply as "the connections between testing and learning" (p.298). While, Gates (1995) claimed, "the washback is the influence of testing on teaching and learning" (p.101). Shohamy (1992) went further when he defined washback as "the utilization of external language test to affect and drive foreign language learning...this phenomenon is the result of the strong authority of external testing and measure impact it has on the lives of test takers". (p.513)

Messick (1996) provided an even more elaborate definition of washback when he wrote "a washback, a concept prominent in applied linguistics, refers to the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning" (p.241)

### **1.7.6. Interactiveness**

Interactiveness is another important criterion in language tests. Bachman and Palmer (1996) defined it as "the extent and type of involvement of the test taker's individual characteristics in accomplishing a test task" (p.25). More specifically, individual characteristics i.e. the test taker's language ability, topical knowledge, and affective schema, which are engaged in a test, may influence the candidates' performance in a test.

### **1.7.7. Impact**

The impact is the last criterion of a test. It refers to the effect of a test on the individual test takers, other stakeholders (such as teachers, parents...), educational systems, or society. (Richards & Schmidt, 2002, p. 248).

Bachman and Palmer (1996) stated that another quality of tests is their impact on society and educational systems and upon the individuals within those systems. Thus, the quality of tests may have an impact at two levels either micro or macro. The micro level that is affected by the particular test use such as the learners. The macro level refers to the educational system or society. (pp. 29-30)

## **1.8. Competency Based Assessment**

Competency based assessment refers to an important process of measuring the students' competencies in particular domain. Therefore, it aims at determining whether students can perform a task or an activity in addition how well they can do them. Consequently, it is quietly significant to understand what is competency based assessment as well as its principles.

### **1.8.1. Definition of the Competency Based Assessment**

According to Wolf (2000), competency based assessment is a form of evaluation that derived from the specification of a set of outcomes. He added there are three main elements of the competency based assessment:

- Competency based assessment emphasises on the outcomes.
- Those outcomes should be specific to the point where they are clear and transparent. Consequently, the assessors, assesses, and the third parties should be able to understand what is being assessed and what should be achieved.
- The decoupling of assessment from particular institutions or learning program.

(pp.1-2)

Dubois and Rothwell (2000) defined competency based assessment as "an evaluation of an individual performer's competence in comparison to some performance requirement or expectation". (p.7)

### **1.8.2. Principles of Competency Based Assessment**

Successful assessment reflects on various principles. Those principles are validity, reliability, flexibility, and fairness.

#### **1.8.2.1. Validity**

Hager, Gonzi, and Athamason (1994) claimed that assessment is said to be valid, if the assessors are fully aware of what is to be assessed, evidence is collected from tasks that are clearly related to what is to be assessed, also the amount of sampling of different evidence to demonstrate the performance is another factor that make the assessment valid. (p.6)

### **1.8.2.2. Reliability**

Hagar, et al. (1994) stated that assessment is said to be reliable, when it is applied and interpreted consistently from one student to another and from one context to another. In addition, (Shepard, 2006, p. 642) claimed that reliability refers to "the process of consistency requirements such as inter-judge agreements for scoring rubrics". (As cited in McMillan, 2013, p.117)

### **1.8.2.3. Flexibility**

Another principle of competency based assessment is flexibility. According to Hagar, et al. (1994) assessment is flexible when it is interpreted successfully and adapted to a range of training modes as well as the students' needs. That is to say, assessment shall be flexible by making reasonable adjustments to assessment techniques students.

### **1.8.2.4. Fairness**

The last principle of competency based assessment is fairness. Hagar, et al. (1994) claimed that assessment is fair when it does not affect the learners negatively. That is to say, fairness is happened when all the learners understand and assess in the same way. For example, when the whole group of the learners understand what is expected of them and what form of assessment will be used.

## **1.9. Reasons for Testing**

Ur (1996) stated nine reasons for testing whereby tests may be used as a means to:

1. Give the teacher information about where the students are now, to help decide what to teach next.

2. Give the student information about what they know, so that they also have an awareness of what they need to learn or review.
3. Assess for some purpose external to current teaching (a final grade for the course, selection).
4. Motivate students to learn or review specific material.
5. Get a noisy class to keep quiet and concentrate.
6. Provide a clear indication that the class has reached a 'station' in learning, such as the end of a unit, thus contributing to a sense of structure in the course as a whole.
7. Get students to make an effort (in doing the test itself, which is likely to lead to better results and a feeling of satisfaction).
8. Give students tasks which themselves may actually provide useful review or practice, as well as testing.
9. Provide students with a sense of achievement and progress in their learning.

(p. 34)

### **1.10. Designing a Test**

Ur (1996) stated that there are three main stages in designing a test; the first stage of designing a test is "the preparation stage". In this stage, as its name denotes, the teachers prepare for doing a test. Therefore, they formulate the learning objectives. In addition, they determine the purposes of testing. Moreover, the teachers list in writing all the materials that the test will cover. Consequently, this stage is helpful for the teachers since it is a guide, plan, and outline for making a test.

The second stage in designing a test is called "performance stage". In the performance stage, the teacher administers the test to a group of students. The last stage is called "feedback of the test".

During this stage, the teachers look at how the test was done. Then, they provide the students with the appropriate feedback and comment.

### 1.11. Stages of Tests Development

Test development refers to a linear process of creating and using the tests. Developing a test starts when the test designers decide that the new test is necessary (Bachaman & Palmer, 1996). According to the association of language testers in Europe (ALTE), test development organized into three main phases: planning, design, and try-out.



**Figure 2: The Test Development Process**

#### 1.11.1. Planning

The planning is the first stage in the test development. In this stage, the sponsor of the test and a wide range of stakeholders (including ministries and government bodies, publishers, schools, parents, experts, employers, educational institutions and administrative centres) are the responsible for providing information about the test requirements, needs, purposes, and personal knowledge i.e. characteristics of the test takers and required ability level. (ALTE, 2011)

#### 1.11.2. Design

The data gathered in the planning stage are considered as the starting points of the design stage. During this phase, the test components are described in details. In other words, the structures and the content of the test are specified. (ALTE, 2011)

### **1.11.3. Try-out**

The last step in developing the test is the try-out. According to Bachman and Palmer (1996), try-out phase involves administering the test for the purpose of collecting information about the usefulness of the test itself, and improving the test and testing procedures. In other words, tests are often tried out before the actual level to check their effectiveness. Moreover, feedback obtained from the try-out phase, consists of some editing, involving returning to the design one and rethinking some of the components in the design stage.

## **1.12. Payoffs and Pitfalls of Test**

### **1.12.1. Payoffs**

- Well-constructed tests sample student's knowledge with efficiency and reliability. The test given can determine what many students know in a brief period.
- The repeated use of test will provide a means of comparison between different student groups or the same group over time.

### **1.12.2. Pitfalls**

- Tests lack flexibility because they are usually designed by the teachers. The content of a test is predetermined.
- Tests can be expensive. Costs associated with purchasing tests and processing results must be considered during assessment planning.

## **Conclusion**

In this chapter, the major concern related to language testing was discussed. First of all, some definitions of tests and its relation to assessment and teaching were reviewed. Moreover, the different types of tests were discussed. The chapter also stressed the importance of tests.

Furthermore, criteria of good testing were discussed. In addition, it referred to the competency based assessment and its principles as well. Finally, the main reasons, stages for designing, payoffs, and pitfalls of tests were presented.

## **Chapter Two: The Achievement Test Evaluation**

### Introduction

#### 2.1. The Historical Development of Achievement Test

#### 2.2. Definition of Achievement Test

#### 2.3. Types of Achievement Test

##### 2.3.1. Teacher Made Test

##### 2.3.2. Standardized Test

##### 2.3.3. Performance Test

##### 2.3.4. Final Achievement Test

##### 2.3.5. Progress Tests

#### 2.4. Functions of Achievement Test

#### 2.5. Principles of Achievement Test

#### 2.6. Planning the Achievement Test

#### 2.7. Definition of Evaluation

#### 2.8. Types of Evaluation

##### 2.8.1. Formative Evaluation

##### 2.8.2. Summative Evaluation

##### 2.8.3. Illuminative Evaluation

##### 2.8.4. Process Evaluation

##### 2.8.5. Product Evaluation

2.8.6. Predictive Evaluation Vs Retrospective Evaluation

2.8.7. Impact Evaluation

2.9. Purposes of Evaluation

2.10. Participants in the Evaluation Process

2.11. The Role of Classroom Testing

2.11.1. Definition of Course Objective

2.11.2. Stimulation of Student Progress

2.11.3. Evaluation of Classroom Achievement

2.12. Test Evaluation

Conclusion

## **Introduction**

The achievement test is administered most frequently in language programmes than any other types of test. It plays a prominent role in all types of instructional programmes. The achievement test is considered as the most widely method that the teachers use in order to assess the students' achievement in the classroom. The achievement test evaluation is the main concern of this chapter. It begins by giving a brief history of the achievement test, defining this type of test, as well as indicating its types and principles. In addition, we highlight the major functions of achievement tests. Besides, this chapter sheds light on the main steps in the planning of the achievement tests. Moreover, the term evaluation is defined in this chapter. Furthermore, we discuss the different types of evaluation. Besides, the purpose of evaluation and the participants in the evaluation are mentioned in the second chapter. At the end of this chapter, we referred to the role of the classroom testing and the importance of test evaluation as well.

### **2.1. The Historical Development of Achievement Test**

The achievement tests were first developed in 1895 by Race. He defined a spelling test of 50 words with alternate forms to 16000 students in grades four through eight across the country. At that time, Rice aimed to develop test in arithmetic, language, and scientific approach to the assessment of students' knowledge. However, the main contribution was his objective. In fact, the achievement tests were developed across time.

In the first decade of the twentieth century, there were various types of achievement tests developed. For instance, test batteries were developed in 1920, and Stanford achievement test at the elementary level in 1923. Moreover, in 1940, more attention has been given to testing in broad areas i.e. humanities and natural sciences. Furthermore, the attention has been shifted toward evaluation of work study-skills, comprehension, and understanding. As a result, standardized tests were developed that were keyed to particular test book in 1970's.

In the era of 1990's, the main concern of achievement tests was based on the assessment of learning achievement that is built into the instructional process. However, the concern had been directed toward the intrinsic nature of achievement test itself. Besides, the approach of computer-adaptive testing was based on a concept of a continuum of learning and was particular child fits on that continuum. Therefore, the child experience with testing is one of the success rather than the failure. Additionally, the use of alternative assessment tools has taken a front-row seat. Thus, performance of based assessment approach includes testing methods that require students to create an answer or product that demonstrates knowledge or skills.

In the later 19<sup>th</sup> and early 20<sup>th</sup> centuries, the multiple choice testing replaced the alternative forms due to the difficulties with standardization and their use with large number of people. Consequently, in order to accept the shift in the conceptualization of the assessment of achievement it is necessary to understand the nature of tests which fall under the domain of achievement, psychometric underpinnings of achievement tests, the basis for criterion referenced as opposed to norm-referenced measurement, and special issues which rise when achievement tests are used for particular purposes (Katz & Slomka, 1990, pp 150-151)

## **2.2. Definition of Achievement Test**

Achievement tests have been defined by different scholars and researchers as follow:

In the dictionary of language testing (1999), achievement test is defined as a tool that is designed to measure what a person has learned within a given time. It is based on a clear and public indication of the instruction that has been given. (p.2)

Richards and Schmidt (2002) defined achievement test as a test that is designed to measure how much of a language learners have successfully learned with specific reference to a particular course, textbook, or programme of instruction. Thus, a type of criterion-referenced test. An achievement test is typically given at the end of a course...its results are often used to make

advancement or graduation decisions regarding learners or judge the effectiveness of a programme, which may lead to curriculum changes. (p.7)

According to Henning (1987), achievement tests are instruments that enable us to measure the extent of learning in a prescribed content domain, often in accordance with explicitly stated objectives of a learning program. (p.6)

Bachman (1990) claimed that achievement tests or attainment tests referred to a short tests or quizzes that the classroom teachers give to the students at the end of each instructional unit or group of unit to provide additional information on student progress. Such testing is often facilitated by the inclusion, in published materials, and of unit tests. These tests are based on the content of the course. (pp. 60-61)

In more practical and pedagogical terms, Brown (1994) defined achievement tests as tests that are limited to particular material covered in a curriculum within a particular time frame. In other words, tests are designed primarily to measure individual progress rather than as a means of motivating or reinforcing language. Ideally, achievement tests are rarely constructed by classroom teachers for a particular class (p. 259). Besides, he stated that an achievement test is related directly to classroom lessons, units, or even a total curriculum. Moreover, they are limited to particular material addressed in a curriculum within a particular period and are offered after a course has focused on the objectives in question. In addition, achievement tests are often summative because they are administered at the end of a unit or term of study. (Brown, 2004, p. 48)

McNamara (2000) claimed that achievement tests are associated with the process of instruction. They accumulate evidence during, or at the end of a course of study in order to see whether and where progress has been made in terms of the goals of learning. He added, achievement tests should support the teaching to which they relate. (p. 6)

Keith (2001) suggested that the achievement tests are concerned with how well a learner has done in relation to a particular course or programme. They are usually come at the end of programmes, and are deliberately based on the content cover in it (p. 292). Whereas, Harmer (2007) claimed that achievement test is done at the end of a term to see how well students have learnt everything. He added, this type of test include a variety of test types and measure the students' abilities in all four skills, as well as their knowledge of grammar and vocabulary. (p.166)

In brief, an achievement test, also referred to as attainment or summative test, is devised to measure how much of a language someone has learned with reference to a particular course of study or programme of instruction. For instance, at the end of the year tests designed to show mastery of the language. An achievement test might be a listening comprehension test based on a particular set of situational dialogues in a textbook. The test has two fold-objectives; the first is to help the teachers judge the success of their teaching. Then, the second is to identify the weaknesses of the learners.

### **2.3. Types of Achievement Test**

Achievement tests generally share some common goals, measuring what students know and can do, improving instruction, and helping students to achieve higher standards. This means that not all the achievement tests are of the same kinds. In other words, there are various types of achievement tests.

Different researchers classified the kinds of achievement tests in slightly different ways. According to Linderman (1967) (as cited in Mukhaja, 2017, pp. 153-160), academic achievement tests are three types:

### **2.3.1. Teacher Made Test**

Teacher made test, as its name denotes, is made by the teachers for a local use. This type of achievement test is prepared only to measure the achievement of specific instructional objectives related to a particular unit of work. They are constructed by teachers for using largely within their classroom.

### **2.3.2. Standardized Test**

Standardized test is a test constructed by the test specialists. It called a standardized test because the designers have administered and scored it under standard and uniform testing conditions. In this type of achievement tests, items are fixed and not modifiable. Hence, a standardized test is a test that has a standard procedure for administration, scoring, and interpretations.

### **2.3.3. Performance Tests**

Performance tests are those tests that require the examiners to perform a task rather than answer some questions. Such tests prohibit the use of language items.

On the other hand, other language tests distinguished between two types of achievement tests, which are the final or standardized achievement tests, and progress tests.

### **2.3.4. Final Achievement Tests**

Final achievement tests are those tests that administered at the end of a course of study. In other words, final achievement test, as its name denotes, is designed at the end of the course of the study. In addition, it can be written, or administered by the ministries of education, official examining boards, or by members of teaching institutions. This type of achievement test is also known as standardized achievement test. (Hughes, 2003, p.13)

### **2.3.5. Progress Tests**

Progress test, as another type of achievement tests, defined by Hughes (2003) as a test which refers to the measures that inform us about the students' progress. (p.13)

### **2.4. Functions of Achievement Test**

To teach without evaluating the extent of learning would be foolish. Thus, tests are used to evaluate and measure the achievement of learners. The achievement type of tests is often constructed with tests that measure aptitude, a more general, and stable cognitive trait. As a result, there are many functions of achievement tests.

The major function of classroom achievement test is to measure the students' achievement as well as to contribute in the evaluation of the educational progress and attainments. That is to say, the students' knowledge and performance is more important than their scores of the tests. Besides, achievement tests help the teachers as well as the instructors to assign meaningful and accurate grades. As a result, these grades are meant to be a compressive measurement of students' achievement. In addition, they are reported to students and their parents to indicate how effective their efforts have been, and may influence the opportunities for future education and employment. Thus, it is necessary for the teachers and the instructors to take seriously their responsibilities for assigning grades.

Another function of classroom achievement tests is to motivate and direct students learning. That is to say, students and teachers as well support the position that students do tend to study harder when they expect an exam than when they do not. Moreover, students emphasize in their studying those things on which they expect to be tested. Achievement tests may motivate them in the sense that they know what the test will require, and if the test does a good job for measuring the achievement of the essential course objectives. Hence, they will better do in the test.

Classroom achievement tests do serve other useful educational functions. That is to say, the process of building them must cause instructors to think carefully about the objectives of instruction in a course. Subsequently, it should cause them to define their objectives operationally (Ebel & Frisbie, 1991, pp. 30-31)

### **2.5. Principles of Achievement Test**

The achievement test is a tool for the teachers used for assessing the students in the educational situations. The extent to which the achievement tests contribute to improve learning and instruction is largely determined by many principles underlying their development and use.

Firstly, the achievement tests must measure the learning outcomes that are in harmony with the instructional objectives. Besides, they seek to encourage the learners to focus on a limited aspect of the course content or to direct their attention to all of the important areas. Consequently, they are designed to improve the students learning. Additionally, achievement tests are always a matter of sampling. In other words, they used to measure a representative sample of the learning outcomes and the subject matter in the instruction. Furthermore, achievement tests include the types' items that are correspondent for assessing the desired outcomes. Finally, they must be designed in order to fit a particular uses to be made of result (Gronlund, 1977, pp.7-14)

### **2.6. Planning the Achievement Test**

The planning of an achievement test can take many forms, but both professional test makers and classroom teachers have found the following series of steps to be most useful.

- Determine the purpose of the test.
- Identify the learning outcomes to be measured by the test.
- Define the learning outcomes in term of specific, observable behavior.
- Outline the subject matter to be measured by the test.

- Prepare a table of specifications.
- Use the table of specifications as a basis for preparing tests.

(Gronlund, 1977, p.18)

In brief, from the above list, it is obvious that the major consideration in the achievement test planning is to determine what to be measured, and to describe it in such precise terms that the test items can be constructed.

## **2.7. Definition of Evaluation**

The process of evaluation is viewed as an important issue in the educational world. It has been defined by different researchers and scholars.

Banks (1977) defined evaluation as a technical aspect of instruction and a part of the teaching and learning process that gives continuous feedback of data to keep the system in adjustment. Besides, Weiss (1972) stated, "evaluation is the systematic gathering of information for the purpose of making decisions" (p.5). Another definition is that of Nevo (1977), he claimed that "evaluation refers to the process of delineating, obtaining, and providing information on the merit of goals, designs, implementation and outcomes of educational activities, and should help to improve an educational product during the process of its development, and/or demonstrate the merit of the final product when its development is completed" (p.127).

For Lynch (1996), evaluation is the systematic attempt to gather information in order to make judgments or decisions. As such evaluative information can be both qualitative and quantitative in form, can be gathered through different methods such as observation or the administration of pencil and paper tests (p.2). Whereas, Gronlund (1981) defined evaluation as a systematic process of determining the extent to which instructional objective are achieved by pupils (as cited in Nunan, 1992, p.184). Furthermore, Nunan (1992) stated that evaluation refers to a wider

range of processes that may or may not include assessment data. In other words, it involves not only assembling information but also interpreting those information-making value judgments. (p.185)

Rea-Dickens and Germaine (1992) provided another definition of evaluation. They claimed that evaluation is the principal and systematic collection of information for the purposes of decision-making. In addition to that, Anderson (1986, p.5) claimed, "Evaluation is the process of seeking to establish the value of something for some purpose..."(As cited in Brown & Rodgers, 2002, p. 227).

Richards and Schmidt (2002) stated that evaluation in general, is the systematic gathering of information for the purposes of decision-making. It may use quantitative methods (e.g. tests), qualitative methods (e.g. observations, ratings), and value judgments. They added, in language planning, evaluation frequently involves gathering information on patterns of language use, language ability, and attitudes towards language. Then, in language programme evaluation, evaluation is related to decisions about the quality of the programme itself and decisions about individuals in the programme. The evaluation of programmes may involve the study of curriculum, objectives, materials, and tests or grading systems. The evaluation of individuals involves decisions about entrance to programmes, placement, progress, and achievement. In evaluating both programmes and individuals, tests and other measures are frequently used. (p.188)

From the above definitions, different researchers and scholars argued that evaluation is a continuous process that is designed to collect data for the guidance of making decision on a particular plan. In addition, it is used to measure whether everything is appropriate or there is a need for change or modification for the reason of improvement.

## **2.8. Types of Evaluation**

There are various types of evaluation depending on what is being evaluated and the purpose of evaluation. Some of them include the following:

### **2.8.1. Formative Evaluation**

Formative evaluation is one of the main types of evaluation. It is an ongoing in course evaluation and contributes to the modification of the course, if necessary (Richard, 2001). In addition to that, (Scriven, 1967) (as cited in Ebel & Frisbie, 1991, p.24) stated that formative evaluation is used for the purpose of assessing a programme and checking whether it is feasible, appropriate, and acceptable before it is implemented. Then, it helps to provide feedback to the teacher and to the student about how things are going. Thus, it assists to refine and improve the programme. Moreover, Richards and Schmidt (2002) claimed that formative evaluation refers to the process of providing information to curriculum programme and materials. In other words, formative evaluation is carried out as a part of the process of programme development in order to find out:

- ✓ What is working well; and what is not.
- ✓ What problems need to be addressed?
- ✓ In addition, the implementation of the programme. (p.209)

Nunan (1992) defined this type of evaluation as the process of providing information to curriculum develop during the development of a curriculum or programme in order to improve it. Moreover, formative evaluation is used in syllabus design, the development of language teaching programmes, and materials. (pp. 191-193)

In short, formative evaluation is a process of evaluation that is made from time to time in the case of an instructional programme and from one stage to the other. It does not provide a totalitarian impression of the quality either of the instructional programmes, the techniques and methods, materials or media.

### **2.8.2. Summative Evaluation**

Another type of evaluation is called summative evaluation. It refers to the end of the course evaluation. Then, summative evaluation helps the teachers to find out if the objectives are really

achieved and if the method and materials have been useful for the purpose to achieve the objective. It aims to provide for a going modification. (Richard, 2001)

Scriven (1967) claimed that summative form of evaluation is conducted at the end of programme design. It gives information about programme which helps in deciding whether to continue or end this programme. It is used to assess whether the result of the programme meet the stated goals. (As cited in Ebel & Frisbie, 1991, p.24)

Richards and Schmidt (2002) claimed that summative evaluation is the process of providing information to decision-makers, after the programme is completed about whether or not the programme was effective and successful. In other words, it is done at the end of the programme. In addition, summative evaluation provides information about the overall effectiveness of the programme. Thus, it is concerned with determining the effectiveness of a programme has been implemented (p.209). Nunan (1992) supported this idea; he stated that summative evaluation is the process of providing information to decision makers, after the programme is completed about whether or not the programme was effective and successful. (p.193)

In brief, summative evaluation is conducted at the end of an instructional segment to determine if learning is sufficiently completed to warrant moving the learner to the next segment of instruction.

### **2.8.3. Illuminative Evaluation**

Illuminative evaluation is an approach to evaluation that seeks to find out how different aspects of a course work or how a course is being implemented and the teaching /learning processes that it creates. It seeks to provide a deeper understanding of the processes of teaching and learning that occur in a programme without necessarily seeking to change the course in any way as a result. (Richards & Schmidt, 2002, p.247)

Parlett and Hamilton (1977, p.10) stated that the primary concern of illuminative evaluation is with the description and the interpretation rather than measurement and prediction.

Its aims are:

- ❖ To study the innovatory programme how it operates how it is influenced by the various school situations in which it is applied; whose those directly concerned regard as its advantages and disadvantages and how students' intellectual tasks and academic experiences are most affected.
- ❖ To discover and document what it is like to be participating in the scheme, whether as a teacher and/ or pupil.
- ❖ To discern and discuss the innovations' most significant features, recurring concomitant and critical processes (As cited in Stufflebean, Madaus, & Kellaghan, 2000, p.30)

That is to say, illuminative evaluation refers to the evaluation that seeks to find out how different aspects of the programme work. Otherwise, a deeper understanding of the teaching and learning processes that occur in the programme necessarily seeking to change the course in any way as a result.

#### **2.8.4. Process Evaluation**

Another type of evaluation is known as process evaluation. It is defined by Long (1984) as "the systematic observation of classroom behavior with reference to the theory of second language development which underlines the program being evaluated". (p.51)

In more details, process evaluation involves the collection of information to describe what programme includes over time. This type of evaluation can occur only once, periodically throughout the duration of the implementation of a programme, or continuously; i.e. it takes place

during the implementation of a programme. Process evaluation focuses on how programme strategies were implemented as planned. In other words, it is used to determine why an established programme has changed overtime. Moreover, it aims to identify the effectiveness of a programme under application. Furthermore, process evaluation focuses on how a programme actually functions and assesses the materials and activities.

### **2.8.5. Product Evaluation**

According to Long (1984) "most product evaluations are product oriented. That is, they focus on what program produces, chiefly in terms of students learning, but sometimes also in terms of changes, it brings about in teachers' and students' attitudes, students' self-concept, related intellectual skills, and the like. Thus, most product evaluations set out to answer one or both of the following questions:

- ✓ Does the program Y work?
- ✓ Does program X work better than program Y? " (p.54)

In brief, product evaluation is an outcome testing evaluation. It includes determining and examining the general and specific outcomes of the programme. Moreover, it includes measuring anticipated outcomes, attempting to identify unanticipated outcomes, and assessing the merit of the programme. In addition to that, the main focus of product evaluations is to measure, interpret, and judge the enticement of a programme. It is mainly concerned with whether the programme has achieved the designed objectives and what changes are required to make the programme effective.

### **2.8.6. Predictive Evaluation Vs Retrospective Evaluation**

Ellis (1997) distinguished between predictive and retrospective evaluation. He claimed that predictive evaluation is designed to make decision regarding what materials to use. As a result, teachers are required to carry out a predictive evaluation of the material available to them in order

to determine which are best suited to their purposes. However, retrospective evaluation is designed to examine materials that have actually been used. It can be impressionistic as well as empirical. Consequently, teachers have used the materials that they may feel the need to undertake a further evaluation to determine whether the materials have worked for them.

In other words, predictive evaluation refers to the evaluation of teaching materials that may be done before they are used in the classroom in order to determine whether they suit the need of particular group of learners. In contrast, retrospective evaluation is the evaluation of teaching materials that may be done after the materials have been done in the classroom in order to evaluate their effectiveness and efficiency, and the teachers as well as the learners' attitudes towards them.

### **2.8.7. Impact Evaluation**

Rossi, Freeman, and Lipsey (2004) stated that impact evaluation refers to the types of evaluation that assesses the programme effectiveness in achieving its ultimate goals. In other words, impact evaluation gauges the extent to which a programme produces the intended improvements in the social conditions it addresses. In addition to that, this type of evaluation seeks:

- To assess whether desired programme outcomes were attained.
- To assess whether the programme was effective in producing change in the social conditions targeted.
- To assess whether programme impact included unintended side effects. (p.70)

Thus, impact evaluation is essential when there is an interest in determining if a programme is effective in its efforts to ameliorate a target problem, comparing the effectiveness of different programmes, or in testing the utility of new effort to address a particular community problem.

## 2.9. Purposes of Evaluation

Weir and Roberts (1994) distinguished between two major purposes of evaluation:

- Accountability which is an important purpose of evaluation. In the sense that, evaluation usually examines the effects of a programme or project at a significant end point of an educational cycle and it is usually conducted for the benefit of an external audience or decision maker. That is to say, accountability is to demonstrate that the students are learning to the standards expected of them and /or that the programme of instruction is working the way it should.
- Another important purpose of evaluation is the programme development, i.e. evaluation is designed to improve the quality of a programme as it is being implemented. It may involve staffs who are in the programme as well as others who are not and may have a teacher development focus. (p.5)

In addition to that, evaluation is used to make placement, advancement, or related decisions about students' status in programme, course, or unit within a course. Consequently, it often relies heavily on language test results and can involve classroom teachers or other school or district professionals. Yet, another purpose of evaluation is to guide classroom instruction and enhance student learning on a day-to-day basis. (Carter & Nunan, 2001, pp.146-147)

## 2.10. Participants in the Evaluation Process

According to Richards (2001), there are two types of participants involved in the evaluation process that are the insiders and the outsiders. The insiders consist of the teachers, students, and anyone else closely involved in improving, applying, and implementing the programme. That is to say, the insiders comprise those persons who are part of the context of teaching and learning. For example, formative evaluation is often carried out by the teachers. Thus, they can monitor the development of the course. In addition, they examine the difficulties that may encounter in their

teaching process. Moreover, they are examining the degree in which the materials are effective. On the other hand, summative evaluation is closely related to the students who will give feedback on the course, and what they need to obtain language proficiency.

Richards added that the outsiders include others who are not part of the programme, i.e. those who are not directly concerned with the context of teaching and learning. For instance, consultants, inspectors, or administrators. Their job is to supplement the teachers' perceptions of what happen in a course with independent observation and opinion. (p.296)

## **2.11. The Role of Classroom Testing**

Testing has become important and prominent in our educational system. Through time testing gains interest. As a result, it has a significant role in the process of teaching and learning. Valette (1977) supported this idea. He claimed that classroom testing has three major functions in the second or foreign language, which are:

### **2.11.1. Definition of Course Objective**

The use of classroom tests in an effective way may help to define the course objective. That is to say, they define the short-term course objectives expected by the teacher. In addition, classroom tests may anticipate the content and the nature of the language-learning programme.

### **2.11.2. Stimulation of Student Progress**

According to Valette (1977), "the test best fulfills its function as part of the learning process if correction performance is immediately confirmed and errors are pointed out" (p.4). In other words, classroom test gives the teachers the opportunities to measure the learners' achievement in the target language, i.e. whether they acquire or master a specific language items.

### **2.11.3. Evaluation of Classroom Achievement**

Most teachers stress that testing on a common and regular basis provides them with valuable information about the difficulties that the learners face. So that, the teacher decides what are the language aspects that require more clarification, explanation, and construct the appropriate activities. Specifically, testing supplies teachers with clues and details related to effectiveness of specific teaching methods. In addition, it helps to make an objective evaluation of the learner's progress individually and in relation to the other classmates. In other words, the process of testing is seen as a bridge building between teaching and learning. Hence, the classroom test is viewed as a mirror for the teacher and the learners to see their reflection clearly. (Valette, 1977, p.4)

### **2.12. Test Evaluation**

Test evaluation plays a crucial role in classroom evaluation from both teachers' and learners' perspectives. For the teachers, classroom test evaluation is viewed as a basis tool. In the sense that, it helps them to decide whether their methods and techniques were effective. In other words, test evaluation is used to measure the success and the failure of the teaching process. Moreover, it enables the teachers to comprehend their pros and strong areas, and recognize their weaknesses. Therefore, it helps them in correcting and improving their weak areas. Additionally, it helps the teachers to discover the students' needs. Further and no less importance, test evaluation helps students in their learning. Test evaluation is made to assess learners' achievement. Besides, it helps to motivate the learners to study, learn more, developing abilities and skills. In addition, test evaluation aids learners in encouraging good study habit. Last but not least, it enables them knowing the results of the progress and getting appropriate feedback.

(Miller, Linn, & Gronlund, 2009, pp. 139-141)

**Conclusion**

In this chapter, the major issues related to achievement test evaluation were discussed. At the beginning, a brief history of achievement test was indicated. In addition, definition of achievement test and its types followed by the principles and functions of the achievement test were presented. In addition to that, different steps of planning the achievement tests were reviewed. Moreover, the chapter provides the definition of the term evaluation. Then, it stressed the types, the purposes, and the participants in the process of evaluation as well. At the end of the second chapter, the role of classroom testing and the test evaluation were analyzed.

The next chapter will explain the research design, and the data analysis of the current research in terms of the participants, procedure, and discuss the results in details.

## **Chapter Three: The Field Works**

Introduction

3.1. Research Design

3.2. Participants

3.2.1. Population

3.2.2. Sampling

3.3. Data Collection Instrument

3.4. Data Collection Procedures

3.4.1. Description of the Questionnaire

3.4.2. Analysis and Interpretation of the Questionnaire

3.5. Discussion of the Results

3.6. Limitations of the Study

3.7. Recommendations and Pedagogical Implications

3.8. Suggestions for Further Research

Conclusion

General Conclusion

References

Appendices

Résumé

الملخص

## **Introduction**

The third chapter in hand is practical in nature. The latter aims to provide a concrete picture about the design and development of achievement tests in secondary school. In order to fulfill this aim, data are collected by using a questionnaire as a tool. This chapter is started by presenting the research design, the participants, and the data collection instrument as well. It is then devoted to the presentation, analysis, and interpretation of those data. Finally, the chapter ends with the discussion of the results, some recommendations, suggestions, and limitations of the study.

### **3.1. Research Design**

The current research is a descriptive study. The aim of this study is to evaluate the design and the development of achievement tests in English. In order to fulfill this aim, test the hypothesis, and to answer the research questions a quantitative method is used. In addition, this method is chosen because it is the most suitable one for gathering quick, easy, and direct data. Moreover, quantitative method is mainly related to the explanation and the description of the research problem, provided a specific and numerical data in order to check the validity and reliability of the data collected.

### **3.2. The Participants**

#### **3.2.1. Population**

Generally, population refers to a group of people that helps the researchers in gathering data. In the current study, the target population represents the secondary school English language teachers.

### **3.2.2. Sampling**

The questionnaire is designed for the English language secondary school teachers in order to evaluate the design and the development of the achievement tests in secondary school education. It was distributed in the period between May 12<sup>th</sup> and 20<sup>th</sup>, to thirty (30) teachers of English who teach different levels in different secondary schools in the district of ElMilia. From the thirty teachers, only twenty (20) filled and returned the questionnaire. The other teachers, ten (10), who did not answer the questionnaire claimed that they were preparing for the final term exam.

### **3.3. Data Collection Instrument**

Since the aim of this study is to evaluate the overall usefulness of the achievement tests, it is necessary to elicit and collect data from people who have direct contact with the design and the development of the tests. As a result, data will be collected through using a questionnaire that is considered as the most useful instrument in collecting data easily and quickly. Also, the questionnaire is a convenient tool to gather data from a large group of participants.

### **3.4. Data Collection Procedures**

According to Nunan (1992) a questionnaire is "an instrument for collecting data, usually in written form, consisting of open and/ or closed questions and other probes requiring a response from subjects" (p.231). That is to say, it is a tool used by the researcher in order to gather information. It usually administered in a written form and consists of a list of questions.

#### **3.4.1. Description of the Questionnaire**

The questionnaire is designed to collect the necessary data quantitatively in order to reach the identified objectives and to answer the research questions of the present study. The teachers' questionnaire is started by a small introduction that present the aim of the research. The teachers

were kindly requested and asked to answer the questions by ticking the appropriate choice and/ or making the full statements when necessary.

The questionnaire contains twenty-eight (28) questions. Those questions are a combination of numeric, open ended and closed ones. The numeric questions are presented in (Q1 and Q5). The open-ended questions are those required the respondents to write their own words, represented in the questions (Q7, Q17, Q20, and Q28). The closed questions are divided into two parts. In the first part, the questions (Q2, Q3, Q4, Q6, Q8, Q9, Q12, Q14, Q16, Q18, Q19, Q21, and Q23) are required the teachers to choose one answer by ticking in the right box or by giving "yes" or "no" answer. The second part, the questions (Q10, Q11, Q13, Q15, Q22, Q24, Q25, Q26, and Q27) are clarification questions in which the respondents were asked to choose then to give a further explanation and justification for their answers.

The questionnaire is divided into three main sections. The Background Information is the first section. It started from the Q1 to Q6. Those questions were administered in order to gather personal information about the teachers' age, gender, and qualification. In addition to the teachers' experiences in teaching English at the current secondary school then the level they have taught.

The second section represents the major core of the questionnaire is entitled "Test Design and Development". It covers fifteen (15) questions (from Q7 to Q20). The aim of this section is to collect information about teachers' view in designing and developing the achievement tests. In addition, it seeks to identify the materials that the teachers used, the procedures, and the instructions that they followed in designing the achievement tests. The section also included the Q20 that seeks to provide suggestions about designing and scoring the tests.

The section three is entitled "Teachers' Perceptions". It includes eight (8) questions (from Q21 to Q 28). The chief concern of this section is to elicit information related to the teachers' perceptions toward assessing and scoring the tests.

### 3.4.2. Analysis and Interpretation of the Questionnaire

#### Section One: Background Information

Q1: Age

Table 1

*The Teachers' Age*

	Options	N	%
1	[25-30[	4	20
2	[30-35[	10	50
3	[35-40[	3	15
4	[40-45[	1	5
5	[45-50]	2	10
Total	/	20	100

This question aims at knowing the teachers' age. The respondents in the current study are classified into five categories. The results obtained from the above table show that (½) half of the participants (i.e. 50 %) are between 30 and 35 years old. Next, the teachers who are between 25 and 30 years old (with the percentage of 20 %). Then, 15 % represented the teachers who are between 35 and 40 years old. In addition, two teachers (i.e. 10 %) are between 45 and 50 years old. Last, one teacher (i.e. 5 %) is aged between 40 and 45 years old.

Q2: Gender?

- a) Male
- b) Female

Table 2

*Teachers' Gender*

Options	N	%
Male	4	20
Female	16	80
Total	20	100

It is clearly indicated from the above table that the majority of the teachers are females (80 %). While twenty per cent (20 %) represent males. This means that the majority of teachers in our sample are females. Thus, it can be noticed that the number of female teachers in secondary school is higher than that of male.

## Q3 : Qualification :

- a) Licence
- b) Master
- c) Magister
- d) ENS a Graduate of the Teachers High Training School

Table 3

*Teachers' Qualification*

Options	N	%
Licence	8	40
Master	7	35
Magister	0	0
ENS a Graduate of the teachers High Training	5	25

School		
Total	20	100

This question was addressed to know the teachers' highest degree of education. The results show that the majority of the teachers (40 %) have a licence degree followed by those with master degree (35 %). Then, 25 % are graduates of the teachers' training school (ENS). Whereas, no one has a magister degree.

Q4: How long have you been working as a teacher?

- a) Less than five years
- b) Between five and ten years
- c) More than ten years

Table 4

*Teachers' Teaching Experience*

Options	N	%
Less than five years	7	35
Between five and ten years	7	35
More than ten years	6	30
Total	20	100

The question 4 aims at gathering data from different experiences and to explore various opinions. The results show that 6 informants have been teaching English for more than ten years while 7 of them have a teaching experience between five and ten years and the other seven teachers have taught English for less than five years.

Q5: How long have you been teaching English at this secondary school?

(The number of years)

Table 5

*The Experience of Teaching English at the Current Secondary School*

Number of teachers	Number of years	%
5	2	25
3	1	15
2	5	10
2	3	10
1	29	5
1	18	5
1	17	5
1	15	5
1	11	5
1	8	5
1	6	5
1	4	5
20	/	100

The table 5 explores the number of years of teaching English at the current secondary school. The results obtained show that five teachers (i.e. a percentage of 25 %) started working before two years ago. In addition, 15 % of the teachers started teaching this year. Besides, two of the respondents have taught for five years. Furthermore, ten per cent (10 %) of the English language teachers in the target population have worked for three years at the present secondary school.

Moreover, 29, 18, 17, 11, 8, 6, 4, represent the numbers of years of eight teachers who taught English at the current secondary school.

Q6: What levels have you taught?

- a) 1 AS +2 AS
- b) 1 AS +3 AS
- c) 2 AS + 3 AS
- d) 1 AS+ 2 AS+ 3 AS ( All levels)

Table 6

*The Levels Taught by Teachers*

Option	N	%
1 AS +2 AS	1	5
1 AS +3 AS	2	10
2 AS + 3 AS	2	10
1 AS+ 2 AS+ 3 AS (All levels)	15	75
Total	20	100

The aim of this question is to determine the level (s) that the participants were currently teaching. The results obtained from the table 6 show that the majority (75 %) of the teachers taught all the levels. Ten per cent (10 %) of them taught the first and third levels. In addition, two teachers (10 %) taught the second and third levels. The remaining (5 %) of the respondents taught 1<sup>st</sup> and 2<sup>nd</sup> levels.

## Section Two: Test Design and Development

Q7: Can you define the following words: ‘test’, ‘test design’, ‘test development’?

This question was designed to figure out different meaning to the concepts test, test design, and test development. The teachers’ answers can be summarized as follow:

A ‘test’ is a method used by the teachers during the academic term in order to measure the students’ achievement. That is to say, a test is a form of assessment that intended to evaluate the students’ performance, skills, and knowledge.

‘Test design’ refers to the process that describes how the test should be done.

‘Test development’ refers to the process of creating a better assessment.

Q8: Do you use achievement tests to assess your students’ performance?

- a) Yes
- b) No

Table 7

### *The Use of the Achievement Tests in Teaching*

Options	N	%
Yes	20	100
No	0	0
Total	20	100

According to the table 7, all the participants of the target population use the achievement tests as a way to evaluate and assess their students’ performance.

Q9: Do your students' scores in the test reflect their actual level?

- a) Yes
- b) No
- c) Not sure

Table 8

*The Teachers' Perceptions about the Students' Scores*

Options	N	%
Yes	9	45
No	2	10
Not sure	9	45
Total	20	100

This question aimed to investigate the extent to which the scores of the tests reflect the students' actual level. The results show that 45 % of the informants argued that the scores of the tests reflect the actual level of the students. Nine (9) of the participants i.e. 45 % do not sure whether or not the tests 'scores reflect the students' levels. The remaining of the teachers 10 % said that there is no relation between the students' level and the results obtained in the tests.

Q10: Do you test the listening and speaking skills?

- a) Yes
- b) Sometimes
- c) Never

Why ?

Table 9

*The Place of Listening and Speaking skills in Language Testing*

Options	N	%
Yes	4	20
Sometimes	13	65
Never	3	15
Total	20	100

This question is about whether or not the teachers test the listening and speaking skills. The data gathered revealed that the majority of them test those skills from time to time because they claimed that listening and speaking play a significant role in English language learning, they faced some problems such as the lack of means and the limited time to test them. In addition, 20 % of the teachers (4 teachers) answered positively. In other words, they stated that listening and speaking should be tested because those skills are no less important than the other skills i.e. writing and reading. The rest 15 % of the respondents never test their students the listening and speaking skills mainly because those skills are not recommended in the official 'Examination Guide' that corresponds in its format to the BAC exam model.

Q11: How do you feel when you design an achievement test?

- a) Pleased
- b) Stressed

Why ?

Table 10

*The Teachers' Reactions during the Design of the Achievement Tests*

Options	N	%
Pleased	6	30
Stressed	14	70
Total	20	100

The table 10 clearly indicated that the majority of secondary school teachers (70 %) were stressed when they design the achievement tests mainly because they are going to test their teaching methods. In addition, they feel under pressure since they thought that designing an achievement test that suit their students' needs is a difficult matter. In contrast, 30 % of the participants (six teachers) were satisfied during the design of the achievement tests in the sense that they conceive that testing is a needed operation to assign scores for their students. As a result, their students will get good scores.

Q12: Have you received any training in designing a test?

- a) Yes
- b) No

Table 11

*The Place of Language Testing Design in the Educational Training*

Options	N	%
Yes	13	65
No	7	35
Total	20	100

The data gathered with regard to this question as illustrated in the above table has shown that 65 % of the teachers have received training in designing a test. While, only 35 % of the teachers did not trained in designing the tests.

- If yes, how it was?
  - a) Theoretical
  - b) Practical
  - c) Both

Table 12

*The Place of Educational Training in Language Testing*

Options	N	%
Theoretical	2	15.35
Practical	0	0
Both	11	84.62
Total	13	100

The results of the previous question show that only 13 teachers have received training in designing the tests. Out of them, 84.62 % of the participants (i.e. 11 teachers) have received both a theoretical and a practical training. Whereas, two of the teachers with the percentage of 15.38 % have received only the theoretical training. No one picked out the option of practical training.

Q13: According to you, designing an achievement test requires a special training:

- a) Strongly Agree
- b) Agree

- c) Disagree
- d) Strongly Disagree

Table 13

*The Teachers' Perception toward the Need for a Special Training in Designing the Achievement Tests*

Options	N	%
Strongly Agree	8	40
Agree	12	60
Disagree	0	0
Strongly Disagree	0	0
Total	20	100

We addressed this question to determine whether the design of the achievement tests requires a special training. Twelve (12) participants with a percentage of 60 % agreed that the designing of tests require a special training because they need to know a lot of information before designing their own tests. Whereas, eight teachers with the percentage of 40 % were strongly agree with the requirement of training because they considered it as a difficult and hard task for them. No teacher was opposed the need for training in designing an achievement test. We concluded that all the respondents have agreed with the necessity of training.

Q14: Do you rely on the textbook activities as a source for designing your tests?

- a) Yes
- b) No

Table 14

*The Use of Textbook Activities in Designing the Test*

Options	N	%
Yes	5	25
No	15	75
Total	20	100

The question 14 was addressed to know whether or not the teachers use the textbook activities in designing their tests. The results obtained from the table 14 have revealed that the majority of English language teachers (75 %) at secondary schools did not use the textbook activities as a source for designing their tests. Yet, 25 % of the participants were depended on the activities of the textbook as a source for design the tests.

Q15: Do you rely on other materials while designing the tests?

a) Yes

b) No

If yes, please specify what these materials are?

Table 15

*The Materials Used in Designing Achievement Tests*

Options	N	%
Yes	18	90
No	2	10

Total	20	100
-------	----	-----

The table 15 explores the materials used by the teachers for designing their achievement tests. The data gathered revealed that the majority of the teachers (90 %) said that they were relying on extra materials when designing tests such as books, internet, bank exams, previous tests, web articles , and grammar books. While, 10 % of the teachers (two teachers) did not use extra materials on designing their tests because they may use textbook activities as a source.

Q16: During the design of achievement tests, you follow:

- a) The predetermined objectives of the course
- b) The teachers' guide
- c) The recommendations of the Ministry of National Education
- d) The inspectors' recommendation
- e) Other

If others, would you please specify them?

Table 16

*The Procedures that the Teachers Rely on when Designing Achievement Tests*

Options	N	%
The predetermined objectives of the course	17	32.69
The teachers' guide	6	11.54
The recommendations of Ministry of National	13	25

Education		
The inspectors' recommendation	14	26.92
Other	2	3.85
Total	52	100

In this question, teachers are allowed to choose more than one option. In order to figure out the instructions that the English language teachers followed when designing achievement tests, the question 16 was addressed. The results show that 32.69 % of the teachers were designing their tests depending on the predetermined objectives of the course. While, 26.92 % of the informants were relied on the inspections' recommendation. In addition, (¼) quarter of the questioned teachers (i.e.25 %) were followed the recommendation provided by the Ministry of National Education. 11.45 % followed their guide. Furthermore, only small fractions of the population (i.e. 3.85 %) were dependent on other instructions the learners' need and level.

Q17: What are the procedures that you follow in order to design achievement tests?

This question sought to know the procedures that the secondary school teachers of English followed in order to design an achievement test. In fact, the teachers' responses included various procedures that can be summarized as follow:

- Planning the tests by choosing the suitable tasks that suit the learners' level and set its objectives.
- Designing the test by collecting the appropriate instructions for making a reliable, valid, and useful test.
- Writing items by selecting the appropriate activities of the test.
- Scoring keys which means evaluation and assessing the tests.

Q18: Do you think that all the teachers of English at the Algerian secondary school should follow the same testing procedures?

- a) Yes
- b) No

Table 17

*The Teachers' Perception about Testing Procedures in the Algerian Secondary School*

Options	N	%
Yes	11	55
No	9	45
Total	20	100

The question 18 was introduced to investigate the teachers' view towards whether or not the Algerian secondary schools teachers of English should follow the same testing procedures. Almost half (45 %) of teachers saw that it is not necessary for the English teachers to obey the same procedures in testing.

Q19: How much do you agree with this statement, the test fulfills the general objective of the course.

- a) Strongly Agree
- b) Agree
- c) Disagree
- d) Strongly Disagree

Table 18

*The Test Fulfillment of the Course Objective*

Options	N	%
Strongly Agree	7	35
Agree	12	60
Disagree	1	5
Strongly Disagree	0	0
Total	20	100

The data gathered with regard to this question as illustrated in the table 18 has shown that the majority of the informants, with the percentages of 60 % agreed that the tests fulfill the general objective of the course. Moreover, seven teachers (i.e. 35 %) were strongly agreed that the tests achieve the objectives of the course. However, only one teacher (i.e. 5 %) disagreed that the tests realize the objectives of the course. No teacher ticked the option of strongly disagree.

Q20: Do you have any suggestions about designing and scoring the test?

Regarding this question, the majority of the teachers have not answered this question mainly because they have no suggestions about scoring and designing the tests. The few teachers who have responded to this question suggested some recommendations that they hope to be included in the design and scoring of tests concerning the process of designing the tests, they suggested:

- Including the listening and speaking skills in testing.
- The questions and activities of the test should be equivalent to the students' level and their abilities.

- Including some other behaviors reflexive to the personality, abilities, and interest of the students when scoring the tests.

### Section Three: Teachers' Perceptions

Q21: Assessing the students' achievement is one of the main concerns of the teacher. As a teacher, how do you feel when you score the tests i.e. assessing the students?

- a) Pleased
- b) Stressed
- c) Anxious
- d) Unstressed

Table 19

#### *The Teachers' Perception toward Scoring Tests*

Options	N	%
Pleased	4	20
Stressed	10	50
Anxious	2	10
Unstressed	2	10
Total	20	100

It is clearly indicated that assessing the students' achievement is one of the main concern of any teacher. The question 21 was addressed to know the teachers' feelings when they score the tests. The results in the table 19 revealed that 20 % of the informants were satisfied when they assess their students. Moreover, half of them, i.e. 50 % were stressed during the process of evaluation and assessing their students 10 % of the respondents said that they feel anxious when

they assess their students. While another 10 % of the teachers feel unstressed. The rest 2 of the informants did not choose any option. We concluded that each teacher feel in a slightly different way from another one when assessing his/ her students.

Q22: How much do you agree with this statement?

Assessing the students' achievement is a challenge for the teacher.

- a. Agree
- b. Strongly agree
- c. Disagree
- d. Strongly disagree

Would you please explain your choice?

Table 20

*The Teachers' Perception toward Assessing Students' Achievement*

Options	N	%
Agree	17	85
Strongly Agree	0	0
Disagree	3	15
Strongly Disagree	0	0
Total	20	100

This question was addressed to determine whether or not the assessment of the students' achievement is a challenge for the teachers. It is clearly indicated from the above table that the majority of English language teachers (i.e. 85 %) consider the assessment as a challenge for them

mainly because they thought it is not easy task that required to suit the learners' levels as well as to determine the success or the failure of their teaching process. Whereas, other teachers (15 %) stated the opposite. They seem that the assessment of the students' achievement is not a challenge because every teacher can easily assess his/ her learners' achievement. No one picked out the strong agree or strong disagree options.

Q23: Does the achievement test appropriate to your students' level?

- a) Yes
- b) No

Table 21

*The Appropriateness of the Achievement Test*

Options	N	%
Yes	18	80
No	2	20
Total	20	100

This question aims at knowing the teachers' perception toward the suitability of achievement tests to learners' level. The obtained results show that the majority of the questioned teachers (80 %) stated that the achievement tests are appropriate to their students' level. Whereas, (20 %) of the informants claimed that the tests are not suitable to the levels of their students.

Q24: To what extent the quality of the test affects the students 'attainment'?

- a) Always
- b) Often

- c) Sometimes
- d) Rarely
- e) Never

Why ?

Table 22

*The Effectiveness of Test on Students' Attainment*

Options	N	%
Always	5	25
Often	11	55
Sometimes	3	15
Rarely	1	5
Never	0	0
Total	20	100

The question 24 was addressed to know whether or not the quality of the tests affects the students' attainments. A quarter ( $\frac{1}{4}$ ) of the questioned teachers stated that the quality of the test always affects the students' grade. In addition, more than half of the participants chose the 'often' or 'sometimes' options (70 %). Only a small portion (5 %) of the sampled population claimed that the quality of the test do not influence the achievement of the students. No one has selected the 'never' option.

Why?

The teachers who stated that the quality of the test always affects the students' achievement justified their answer by the students misunderstanding of the instructions presented in the test. However, those who said that the tests quality is often or sometimes influence the students accomplishments, their justifications were as follow:

- The psychological and health problems.
- The lack of preparation for the test.
- The students' perception to the assessing tools.

While, the only one teacher who has considered that the quality of the test rarely affects the students level claimed that the students do not take the test seriously i.e. they rely on active students to provide them with ready answers.

Q25: Do all students get approximate scores in the test?

- a) Yes
- b) No

Table 23

*The Students' Score in the Test*

Options	N	%
Yes	0	0
No	20	100
Total	20	100

It is clearly indicated from the above table that all the participants i.e. 100 % have answered negatively. That is to say, the students do not get equal scores in the test. As a result, the scores of the test identify the success or the failure of the students.

- If no, would you please explain the main reasons behind your students' failure in the test?

As far as the previous question is concerned, the teachers provided some reasons related to the students' failure in the test, as follow:

- The students have different levels, learning styles, and abilities.
- Lack of motivation to the test as well as to learn.
- Lack of revision for the test.
- Misunderstanding of the tests' instructions.
- Lack of some language aspects as vocabulary and grammar, and language skills specially writing.
- Students low self-confidence.

Q26: According to you, should all the teachers of English at the secondary school follow the same scoring procedures?

a) Yes

b) No

Table 24

*The Teachers' Perception about Following the Same Scoring Procedures*

Options	N	%
Yes	11	55

No	9	45
Total	20	100

The aim of this question is to determine whether necessary or not for the teachers of English at secondary schools to implement the same scoring procedures. The data gathered revealed that more than half of the participants stated that it is necessary for the English language teachers at secondary school to follow the same procedures in scoring the tests in order to help all of them to judge the success of their teaching as well as to get a reliable assessment and to ensure fairness for all the students. While, the rest of the teachers with the percentage of 40 % said the opposite i.e. they are not obliged to follow the same scoring procedures. Because the scoring procedures depend on the students' different levels and every teacher has his or her own method or way in teaching and scoring. Thus, it is quite unfair to score all the students in an equal way.

Q27: Do you think the way of assessing the students in the Algerian secondary schools goes along with the rapid changes of the educational system?

- a) Yes
- b) No
- c) Partly

Table 25

*The Teachers' Perception about the Way of Assessment in the Algerian Secondary School System*

Options	N	%
Yes	6	30
No	2	10
Partly	12	60

Total	20	100
-------	----	-----

This question was addressed to determine whether or not the assessment method in the Algerian secondary school goes with the changes in the education system. The data gathered revealed that as, illustrated in the above table, show that (60 %) of the participants considered the way of assessment slightly goes with the rapid changes in the educational system. Then, (30 %) of the informants argued that assessing the students in the Algerian secondary schools go with the development in the educational system. In contrast, two teachers with the percentage of (10 %) claimed that the method of assessment did not go with the changes in the educational system.

### 3.5. Discussion of the Results

The analysis and interpretation of the responses obtained from the teachers' questionnaire led to a better understanding of the participants' reactions concerning the design and development of achievement tests

First of all, concerning the use of achievement tests, the results show that the teachers in El-Milia Secondary Schools use achievement tests as a way for assessing and measuring the students' attainment and performance. However, through the analysis of the teachers' responses, it was found that they do not follow any operational and procedural way. They do not care whether the achievement tests reflect the actual level of their students. Also, they do not identify as clearly as possible the constructs to be measured and totally neglect the aspect related to the characteristics of students. Moreover, it is clearly indicated from the teachers' answers, they are aware of the importance of training for designing an achievement test in order to design a valid and reliable tests. Therefore, a test that fits the students' levels and needs.

Moreover, regarding test development, it seems that achievement tests designed by those teachers do not resemble to what they teach in class. Instead, it sounds that they rely on readymade tests available on commercial textbooks, internet, previous made tests, and so on in designing their tests. Yet, teachers do not neglect the textbook activities as a source to develop their tests focusing on testing grammar structure and vocabulary items ignoring the major two skills listening and reading. They consider designing an achievement test a challenge for the teachers. This is because they are obliged to follow various instructions and procedures in order to design a suitable test that comprise almost the same typology of tasks that are found in the 'BAC' Exam. Their argument was that they have to follow the same procedure in order to prepare students to answer similar questions before sitting for this exam. This way would at least guarantee better scores. In addition, some factors such as the overcrowded classes and the shortage of resources prevent them to include these skills in their tests.

To sum up the whole, it is very interesting to say that achievement tests for EFL classes in El-Milia region do not fit the requirements of a useful test.

After analyzing the teachers' responses, we conclude that there are various reasons that affect the usefulness of achievement tests. Therefore, it leads to the invalidity, unreliability, and impracticality. Among them there are:

- The lack of required training in designing and developing the achievement tests.
- The lack of the teachers' self-confidence in designing the achievement tests.
- The lack of some language skills mainly listening and speaking.
- Misuse of the CBA i.e. the lack of the integration of the four skills in the tests.

The educational system in the Algerian secondary schools is based on the implementation of CBA that is mainly focused on the integration of the four language skills as well as the development of the learners' competencies. From the teachers' questionnaire, it was surprisingly found that the

achievement tests assess either one and/ or two skills mainly writing and reading. Whereas, they neglect the other skills i.e. listening and speaking. Consequently, the achievement tests do not take into consideration the integration of the four language skills. Thus, we conclude that they are not appropriate to the CBA.

### **3.6. Limitations of the Study**

In fact, we found some problems and limitations during the process of conducting this piece of research.

- The time constraints constituted the great limitation. The short period (approximately five months) is not enough for conducting the research appropriately. As a result of that, it forces us to not cover all the areas of the topic.
- When carrying a study the researcher should collect the references and group them depending on the theme. So, concerning our theme "An Evaluation of Achievement Tests in Secondary School Education", there was a lack of references as far as achievement test is concerned. Thus, we lost much time seeking for the references. That is why we relied on a limited number of books, journal articles, and web sites.
- This study is limited to public secondary schools at the district of El Milia in Jijel. Therefore, conclusions need to be verified by conducting similar studies across other different cities in Algeria.
- This study dealt with a limited number of English language teachers at secondary schools as a sample. Therefore, the results cannot be generalized.
- Another problem concerning the teachers' questionnaire is the fact that though the number of questionnaires was thirty (30), only twenty (20) teachers have returned the copy. In addition, some teachers were not helpful at all; they did not answer all the required questions especially open-ended ones that they just have left blank.

### 3.7. Recommendations and Pedagogical Implications

Based on the results obtained in this study, it would be better to devise the following set of pedagogical implications and recommendations for EFL teachers in the Algerian secondary school about how they should design and develop the achievement tests:

- Teachers should consider some essential test qualities and create a balance between them i.e. validity, reliability, practicality, authenticity, wash back, interactiveness, and impact.
- The teachers should bear in mind the students characteristics as well as their needs, when designing the tests.
- It is recommended for the teachers to include listening and oral activities in the tests.
- It is worthwhile to vary the tests' activities and questions according to the students' levels. For instance, average tests to low achievers and more difficult tests to high students.
- It is recommended for the inspectors and the ministry of national education to provide the teachers with the necessary training, sufficient resources, and time to design and develop the achievement tests.
- Teachers should design a test that supports, enables, and helps the students to perform well.
- It is necessary to raise the teachers' awareness about testing and scoring procedures.
- Teachers should be aware of the significance of integrating the four skills as well as language aspects while they design the achievement tests.
- Teachers should strive that the test instructions are as clear as possible and sufficient.

### **3.8. Suggestions for Further Research**

On the basis of the remarks yielded from the findings of our research, it is noteworthy to empower future researchers with the following suggestions:

- It is suggested to replicate this study and use other types of data collection procedures such as checklist and interview.
- Since the present study is the first one conducted to evaluate the design and development of achievement tests in the Algerian secondary school. The results need to be verified with a large group of teachers from different Algerian regions and cities.
- For future researchers should increase the number of participants to see if that would lead to more gains and better results of designing and developing achievement tests by the teachers.

### **Conclusion**

This chapter was devoted to the research design, the instrument, as well as the data collection tool that has been used for the sake of gathering the required information about the design and the development of the achievement tests. This chapter also covers the analysis of the data gathered and discusses the results collected from the secondary school teachers' questionnaire. All the questions were analyzed, and the responses were displayed in tables. Based on the findings of this chapter, our hypothesis were confirmed. At the end of the third chapter, some limitations that we encountered when conducting this piece of research, recommendations, and some suggestions for further research were presented.

## **General Conclusion**

Achievement tests generally share some common goals, measuring what students know and can do, improving instruction, and helping students to achieve higher standards. The present study attempted to evaluate the design and the development of achievement tests in the Algerian secondary schools. It was hypothesized that if achievement tests are well designed and developed, then the latter would fit the requirements of a useful test.

This dissertation is split up into three (3) main chapters: two theoretical and one analytical. The first chapter provided an overview about language testing; through discussing its different definitions, types, and the criteria of a useful test. Then, it moves to draw attention to the competency-based assessment as well as its types. In addition, to the stages of tests development and design.

The second chapter, on the other hand, considered the achievement tests evaluation. More precisely, achievement tests; its historical development, definition, and mentions its various types. Next, the functions, the principles, and the planning of achievement tests were presented. In addition, the second chapter focuses also on evaluation, its types, purposes, and participants on the process of evaluation as well. At the end, the role of classroom testing and test evaluation were indicated.

The third chapter sheds light on the fieldwork of the results obtained from data collected using a teachers' questionnaire. That was designed to 20 secondary school education. The findings of this research revealed that the teachers' questionnaire emphasized of the use of achievement tests to assess the students' attainment and performance. In addition, it was found that the teachers follow no balance between fundamental test qualities such as validity, reliability, and practicality in order to design a useful test. Then, all the language skills are presented in the achievement tests; however, the emphasis is put more on writing and reading. The listening and speaking have not

given a particular section. This is because they follow the typology and content similar to that presented in 'BAC Exams'. Furthermore, the data gathered showed that there are various reasons that affect the usefulness of the achievement test. Therefore, it leads to the unreliability, lack of validity, and impracticality. In addition, it was found that the achievement tests are not appropriate to the CBA.

Finally, this study reveals that the issue of testing is not an easy task. Rather, it is a very systematic process and a well-principled procedural of gathering data about general and specific language abilities. It is an integral part of teaching and a source of collecting useful feedback as how to improve teaching and enhance learning.

### References

- Alderson, J. C., & Buck, G. (1993). Standards in testing: a study of the practice of UK examination boards in EFL/ESL testing. *Language Testing* 10 (1) 1-26. From
- Alderson, J. C., Clapham, C., & wall, D. (1995). *Language test construction and evaluation*. Cambridge, UK: Cambridge University Press.
- Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The learning House: a Journal of Educational Strategies, Issues, and Ideas*.
- Association of Language Testers in Europe (ALTE), (2011). *Manual for Language Test Development and Examining*. Strasbourg: Council of Europe.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2004 ). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford, UK: Oxford University Press.
- Bachman, L. F., & Purpura, J. E. (2008). Language Assessments: Gate-Keepers or Door-Openers? In B, Spolsky and M, H Francis (Eds.), *The Handbook of Educational Linguistics*. (pp.456-468). Oxford, UK: Blackwell Publishing Ltd.
- Banks, J. A., & Clegg, A. A. (1977). *Teaching strategies for the social studies: Inquiry, valuing, and decision-Making*. Addison Wesley Publishing Company.

- Benmostefa, N. (2014). *Reflections upon the baccalaureate tests as a source of and means for innovation and change in ELT in Algeria*. (Unpublished doctorate dissertation). University of Tlemcen, Tlemcen, Algeria.
- Broughton, G., Barnett, J.A., & Greenwood, T. (1980). *Success with English: The penguin course teachers' handbook 1*. Penguin Books. Middlesex.
- Brown, H.D. (1994). *Principles of language learning and teaching*. New Jersey: Prentice-Hall.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York, NY: Longman.
- Brown, H. D., & Rodgers, T.S. (2002). *Doing second language research*. (Vol.7). Oxford: Oxford University Press.
- Bull, V. (Ed.). (2008). *Oxford: Learner's pocket dictionary* (4th ed.). Oxford: Oxford University Press.
- Carter, R. & Nunan, D. (Ed.). (2001). *The Cambridge guide to teaching English to speakers of other languages*. Cambridge: Cambridge University Press.
- Carroll, B. J. (1968). The psychology of language testing. In A. Davies (Ed.), *Language testing symposium: A psycholinguistic approach* (pp. 46-69). London: Oxford University Press.
- Coomber, C., Davidson, P., O'Sullivan, B., & Stoyhoff, S. (Eds.). (2012). *The Cambridge guide to second language assessment*. Cambridge: Cambridge University Press.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (Eds.). (1999). *Dictionary of Language Testing: Studies in Language Testing, Vol 7*. Cambridge: Cambridge University Press UCLES.
- Dubois, D., & Rothwell, W. (2000). *The Competency toolkit*. Canada, US: HRD Press.

- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5<sup>th</sup> ed). Englewood Cliffs, NJ: Prentice Hall.
- Ellis, R. (1997). The empirical evaluation of language teaching materials. *ELT journal*, 51(1), 36-42.
- Gates, S. (1995). Exploiting wash back from standardized test.
- Genesee, F., & Upshur, J.A. (1996). *Classroom based evaluation in second language education*. Cambridge [England]: Cambridge University Press.
- Gillard, P. (2003). *Cambridge advanced learners dictionary*. Cambridge: Cambridge University Press.
- Gronlund, N. E. (1977). *Constructing achievement tests*. Englewood Cliffs, NJ: Prentice-Hall.
- Hagar, P., Gonzi, A & Athamason, J. (1994). General issues about assessment of competence. *Assessment and Evaluation in Higher Education*, 19 (1) 3-16.
- Harmer, J. (2001). *The practice of language teaching* (3<sup>rd</sup> Ed). Essex: Longman.
- Harmer, J. (2007a). *How to teach English*. England: Pearson Education Limited.
- Harmer, J. (2007b). *The practice of English language teaching* (4<sup>th</sup> Ed). England: Pearson Education.
- Hedge, T. (2000). *Teaching and learning in language classroom*. Oxford: Oxford University Press.
- Henning, G. (1987). *A guide to language testing*. Cambridge, MA: Newbury House.
- Hornby, A. S., Ashby, M., & Wehmeier, S. (2000). *Oxford advanced learner's dictionary of current English*. Oxford: Oxford University Press.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Katz, L.J., & Slomka, G.T. (1990). Achievement test. In *Handbook of psychological assessment* (2<sup>nd</sup> Ed). (pp. 149-201).
- Keith, J. (2001). *An introduction to foreign language learning and teaching*. New York, NY: Longman.

- Larsen-Freeman, D. & Long, M. H. (1991). *An introduction to second language research*. London: Longman.
- Long, H.L. (1984). *Process and product in ESL program evaluation*.
- Lynch, B. K. (1996). *Language program evaluation: Theory and practice*. Cambridge University Press.
- McMillan, J.H. (Ed). (2013). *SAGE handbook of research on classroom assessment*. Virginia Commonwealth University.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Messick, S. (1996). *Validity and Wash back in Language Testing*. *Language Testing* 13(3) 241-256.
- Milanovic, M. (2002). *Language examining and test development*. Strasbourg: Council of Europe, Language Policy division.
- Miller, M. D., Linn, R. L., & Gronlund, N.E. (2009). *Measurement and assessment in teaching* (10<sup>th</sup> Ed). Upper Saddle River, NJ: Merrill /Prentice Hall
- Mukhija, J. (2017). *Construction of achievement tests in English*. Vol 3. (pp.153-162). Bikaner: Maharaja Ganga Singh University
- Naoua, M. (2016). *An evaluation of English language testing in the baccalaureate examination*. (Unpublished doctorate dissertation). University of Mentouri Brothers, Constantine, Algeria.
- Neukrug, E. S., Fawcett, R. C. (2004). *Essentials of testing and assessment: A guide to counselors, social workers, and psychologists*.
- Nevo, D. (1977). *A Model for the utilization of formative evaluation in the process of Developing Instructional Materials*.
- Nunan, D. (1992). *Research methods in language learning*. Cambridge: Cambridge University Press.
- Prator, C. H. (1972). *Manual of American English Pronunciation. Diagnostic Passage*. New York: Holt, Rinehart, and Winston.

- Rea-Dickens, P. & Germaine, K. (1992). *Evaluation*. Oxford University Press.
- Richards, J. C., & Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics* (3<sup>rd</sup> Ed). London, UK: Pearson Education Limited.
- Richards, J. C. (2001). *Curriculum development in language teaching*. New York, NY: Cambridge University Press.
- Rossi, P. H., Freeman, H. E., & Lipsey, M.W. (2004). *Evaluation a systematic approach*. London: Library of Congress Cataloging in Publication Data
- Shohamy, E. (1992). *Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning*. *Modern Language Journal* 76(4) 513-52.
- Shohamy, E., Donitsa-Schmidt, S., and Fierman, I. (1996). *Test Impact. Revisited: Wash back Effect over Time*. *Language Testing* 13(3) 298-317.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow: Longman/Pearson
- Spolsky, B. (1989). *Condition for second language learning*. Oxford: Oxford University Press.
- Spolsky, B. (1995). *Language Testing: Art or Science?* Paper Presented at the 4th AILA International Congress. Stuttgart.
- Spolsky, B. (2008). Introduction: Language testing at 25: Maturity and responsibility? *Language Testing*, 25(3), 297–305.
- Stufflebean, D.L., Madaus, G.F., & Kellaghan, T. (2002). *Evaluation models view points on educational and human services evaluation*. (2<sup>nd</sup> ed). New York NY: Kluwer Academic.
- Stufflebean, D.L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and application*. San Francisco: Jossey-Bass.
- Test definition in English (n.d). In Wikipedia. Retrieved April 4, 2019, from <http://fr.wikiwix.com/index.php?action=test>

Test advantages and disadvantages (n.d). Retrieved April 4, 2019, from <https://cms.bsu.edu/-/media/WWW/DepartmentalContent/Effectiveness/pdfs/Wkbk/WBKM12012%20%20Ch%207.pdf>

Ur, P. (1991). *A course in language teaching: Practice and theory*. Cambridge: Cambridge University Press.

Valette, M. R. (1977). *Modern language testing*. New York, UK: Harcourt Brace Jovanovich.

Weir, C., & Roberts, J. (1994). *Evaluation in ELT*. Oxford: Black Well.

Weiss, C. H. (1972). *Evaluation research: Methods for assessing program effectiveness*. Boston: Prentice Hall.

Wolf, A. (2001). Competency based assessment. In J. Raven & J. Stephenson (Eds.), *Competence in the learning society* (pp.453-466). New York: Peter Lang.

Wyk, B. V. (n.d.). Research design and methods. Retrieved from:

[https://www.uwc.ac.za/Students/Postgraduate/Documents/Research\\_and\\_Design\\_I.pdf](https://www.uwc.ac.za/Students/Postgraduate/Documents/Research_and_Design_I.pdf)

## Teacher's Questionnaire

Dear teacher,

This questionnaire is an important part of a research work. It aims to evaluate the designing of achievement test in the Algerian secondary school. You are kindly requested to answer the following questionnaire. Your answers will remain anonymous. We also tend to inform you that there is neither right nor wrong answers. So, please for the sake of reliability, give your answers as sincerely as can be.

Thank you in advance.

Would you, please, tick (✓) the appropriate answer or write the full statements when necessary.

### Section One : Background Information

1. Age

2. Gender

a. Male

b. Female

3. Qualification :

a. Licence

b. Master

c. Magister

d. ENS a Graduate of the Teachers High Training School

4. How long have you been working as a teacher?

a. Less than five years

b. Between five and ten years

c. More than ten years

5. How long have you been teaching English at this secondary school?

(The Number of Years)

6. What levels have you taught?

a. 1 SA +2 AS

b. 1 SA +3 AS

c. 2 SA + 3 AS

d. 1 SA+ 2 AS+ 3 AS ( All levels)

**Section Two: Test Design and Development**

7. Can you define the following words: 'test' , 'test design', 'test development'

.....  
.....  
.....  
.....

8. Do you use achievement tests to assess your students' performance?

a. Yes

b. No

9. Do scores of your students obtain in a test reflect their actual level?

a. Yes

b. No

c. Not sure

10. Do you test the listening and speaking skills?

a. Yes

b. Sometimes

c. Never

Why?.....  
.....  
.....

11. How do you feel when you design an achievement test?

a. Pleased

b. Stressed

Why ?.....  
.....  
.....

12. Have you received any training in designing a test?

a. Yes

b. No

If yes, how it was?

a. Theoretical

b. Practical

c. Both

13. According to you, designing an achievement test requires a special training

a. Strongly agree

b. Agree

c. Disagree

d. Strongly disagree

Justify your answer

.....  
.....  
.....

14. Do you rely on the textbook activities as a source for designing your tests?

a. Yes

b. No

15. Do you rely on other materials while designing the tests?

a. Yes

b. No

If yes, please specify what these materials are?

.....  
.....  
.....

16. During the designing of the achievement test, you follow:

a. The predetermined objectives of the course

b. The teachers' guide

c. The recommendations of Ministry of National Education

d. The inspectors' recommendation

e. Others

If others, would you please specify them?

.....  
.....  
.....

17. What are the procedures that you use in order to design achievement tests?

.....  
.....  
.....  
.....

18. Do you think that all the teachers of English at the Algerian secondary school should follow the same testing procedures?

- a. Yes
- b. No

19. How much do you agree with this statement, the test fulfill the general objective of the course.

- a. Strongly agree
- b. Agree
- c. Disagree
- d. Strongly disagree

20. Do you have any suggestions about designing and scoring the test?

.....  
.....  
.....  
.....

**Section Three: Teachers' Perceptions**

21. Assessing the students' achievement is one of the main concerns of the teacher. As a teacher, how do you feel when you score the tests i.e. assessing the students?

- a. Pleased
- b. Stressed
- c. Anxious
- d. unstressed

22. How much do you agree with this statement?

Assessing the students' achievement is a challenge for the teacher.

- a. Agree
- b. Strongly agree
- c. Disagree
- d. Strongly disagree

Would you please explain your choice?

.....

.....

23. Does the achievement test appropriate to your students' level?

- a. Yes
- b. No

24. To what extent the quality of the test affects the students 'attainment'?

- a. Always

b. Often

c. Sometimes

d. Rarely

e. Never

Why ?.....

.....

.....

25. Do all students get the same score in the test?

a. Yes

b. No

If no, would you please explain the main reasons behind your students' failure in the test?

.....

.....

26. According to you, all the teachers of English at the secondary school should follow the same scoring procedures.

a. Yes

b. No

Why or Why not ?

.....

.....

27. Do you think the way of assessing the students in the Algerian secondary schools goes along with the rapid changes of the educational system?

- a. Yes
- b. No
- c. Partly

Would you please justify your answer?

.....  
.....

28. Do you have anything to add?

.....  
.....  
.....

**Thank you so much. We appreciate your collaboration 😊**

## EVALUATION OF ACHIEVEMENT TESTS

### Résumé

La présente étude vise à évaluer l'utilité globale des tests de performance en anglais dans les écoles secondaires algériennes. En termes précis, il cherche à déterminer si les tests de rendement reflètent les résultats réels des élèves par rapport au contexte du programme et aux objectifs pédagogiques qu'ils ont définis par la suite. Afin de vérifier la validité de l'hypothèse de recherche, selon laquelle les tests de réalisation répondraient aux exigences d'un test utile s'ils étaient bien conçus et développés, une méthode de recherche descriptive a été utilisée pour la collecte et l'interprétation des données. Cette enquête a été réalisée au moyen d'un questionnaire adressé à vingt professeurs d'anglais de différentes écoles secondaires du district d'El-Milia à Jijle. Les résultats obtenus ont révélé que les tests de réussite des classes d'anglais langue étrangère de la région d'El-Milia ne répondaient pas aux critères d'un test utile. En outre, il a été constaté que les tests de performance ne sont pas appropriés pour la CBA. À la fin, l'étude propose des recommandations pédagogiques relatives à la conception, au score et au développement des tests de réussite dans l'enseignement secondaire.

## EVALUATION OF ACHIEVEMENT TESTS

### الملخص

تهدف هذه الدراسة إلى تقييم الفائدة الإجمالية لاختبارات التحصيل في اللغة الإنجليزية في المدارس الثانوية الجزائرية. عبارات دقيقة، يسعى إلى معرفة ما إذا كانت اختبارات التحصيل تعكس الأداء الفعلي للطلاب فيما يتعلق بسياق المنهج وأهدافهم التعليمية اللاحقة. لغرض التحقق من صحة فرضية البحث، التي تنص على أن اختبارات الإنجاز سوف تلبى متطلبات الاختبار المفيد إذا تم تصميمها وتطويرها بشكل جيد، تم استخدام طريقة بحثية وصفية لجمع البيانات وتفسيرها. تم إجراء هذا التحقيق من خلال إدارة استبيان لعشرين مدرساً للغة الإنجليزية في مدارس ثانوية مختلفة في منطقة الميلية في جيجل. كشفت النتائج التي تم الحصول عليها أن اختبارات التحصيل لفصول اللغة الإنجليزية كلغة أجنبية في منطقة الميلية لا تتناسب مع متطلبات الاختبار المفيد. بالإضافة إلى ذلك، تبين أن اختبارات الإنجاز غير مناسبة لـ CBA. في النهاية، تقترح الدراسة بعض التوصيات التربوية ذات الصلة بتصميم، والنتيجة، وتطوير اختبارات التحصيل في التعليم الثانوي.

## **Abstract**

The present study aims at evaluating the overall usefulness of achievement tests of English in the Algerian secondary schools. In precise terms, it seeks to investigate whether achievement tests reflect the students' actual performance in relation to the context of the syllabus and their subsequent instructional objectives. For the purpose of verifying the validity of the research hypothesis, which states that achievement tests would fit the requirements of a useful test if they were well designed and developed, a descriptive research method was employed for data collection and interpretation. This investigation was carried out by the administration of a questionnaire to twenty teachers of English in different secondary schools at the district of El-Milia in Jijle. The results obtained revealed that achievement tests for EFL classes in El-Milia region do not fit the requirements of a useful test. In addition, it was found that the achievement tests are not appropriate to the CBA. At the end, the study proposes some pedagogical recommendations relevant to the design, score, and development of the achievement tests in secondary school education.

***Key words:*** Evaluation, achievement test, test design, test development