

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

Ministry of Higher Education and Scientific Research

University of Abu-Bakr Belkaid, Tlemcen

Faculty of Letters and Languages

Department of English



**AN ASSESSMENT OF THE SPEAKING SKILL AMONG  
EFL LEARNERS**

**Case of Second Year Licence Students at Tlemcen  
University**

*Dissertation Submitted to the Department of English as Partial Fulfilment for the Degree  
of "Magister" in Assessment and Testing*

Presented by:

Miss. El Batoul BOUAB

Supervised by:

Dr. Rahmouna ZIDANE

**Board of Examiners**

Pr. S. BENMOUSSAT	Prof	President	(University of Tlemcen)
Dr. R. ZIDANE	MC 'A'	Supervisor	(University of Tlemcen)
Dr. N. BOUYAKOUB	MC 'A'	Examiner	(University of Tlemcen)
Dr. A. KHALDI	MC 'A'	Examiner	(University of Tlemcen)

Academic Year: 2015-2016

## **DECLARATION OF ORIGINALITY**

I affirm that this work is my own production; there is no kind of plagiarism

Miss. El Batoul BOUAB

A handwritten signature in black ink, consisting of a stylized, cursive script that appears to be the name 'El Batoul BOUAB'.

For my Parents,  
Without whom nothing would have been possible.  
For all those I love.

## **ACKNOWLEDGEMENTS**

First and foremost, all thanks to Allah the Almighty for helping me to complete this work.

I am deeply grateful to my supervisor Dr. Rahmouna ZIDANE for her great effort without which this work would have never been completed. Her timely and insightful comments at every stage in the preparation of my dissertation, and her effort in reading and providing me with valuable comments have facilitated the writing process.

I would like to express my gratitude to the jury members for their worthy comments and advice.

My sincere appreciation goes to the teachers of oral expression with whom I completed the interview and got valuable data.

Special thanks is addressed to all my teachers during my first year, especially Pr. Smail BENMOUSSAT, Dr. Ali BAICHE, Dr. Nawel BENMOSTEFA, and Dr. Abdellatif SAMMOUD for their valuable and encouraging tips and guidance.

I would also like to thank my colleagues for their spirit of assistance and synergy and for their precious morals. We were really an amazing group

I wish to extend my deep thanks to Dr. Yahia ZEGHOUDI, Dr. Amine BELMEKKI, Dr. Radia BENYELLES, and Pr. Hamza BOUBIKRI, for their guidance and help.

## **ABSTRACT**

Speaking lies at the heart of language learning. People, usually, judge the language proficiency through the speaking competence. Speaking a foreign language is the wish the most demanded by many language learners. However, it is illogic to judge a learner's speaking competence only by intuition. A genuine systematic assessment is the only way that allows us to say that this learner is a competent or an incompetent speaker. The present research work aims at investigating the speaking assessment practices within the English department of Tlemcen University, case of second year licence students. To collect data, three research instruments were utilised: a questionnaire addressed to the students, a semi-structured interview with the teachers of the oral expression module, and a test observation. The gathered data were a fertile source of information that answered the enquiry of the researcher concerning what was being measured in a speaking assessment and how this was measured, as well as the difficulties facing students and teachers in learning, instructing, and assessing this skill. The broader purpose of this research work is first to develop effective speaking assessment practices in our university, and second to improve our learners' speaking competence through benefiting from the positive washback of this assessment in promoting their motivation to learn and to make more efforts. It is worth noting that the prime incentive that induced the researcher to conduct such a study was her own experience with this skill. Although she was among the excellent students, she found, and still finds it difficult to sustain a conversation or express herself fluently. Paradoxically, she got good marks (between twelve and fourteen out of twenty) in all the speaking tests she performed!

# TABLE OF CONTENTS

Dedication.....	i
Acknowledgements.....	ii
Abstract.....	iii
Table of Contents.....	iv
List of Acronyms.....	vii
List of Tables.....	viii
List of Bar Graphs.....	ix
List of Pie Charts.....	x
General Introduction.....	1

## CHAPTER ONE: ASSESSING SPEAKING: A LITERATURE REVIEW

1-1 INTRODUCTION.....	4
1-2 DEFINITION OF BASIC CONCEPTS: Evaluation, Assessment, And Testing.....	4
1-3 TYPES OF ASSESSMENT.....	5
1-3-1 SUMMATIVE AND FORMATIVE ASSESSMENT.....	5
1-3-2 FORMAL AND INFORMAL ASSESSMENT.....	6
1-3-2-1 FORMAL ASSESSMENT.....	6
1-3-2-2 INFORMAL ASSESSMENT.....	7
1-3-3 NORM-REFERENCED AND CRITERION-REFERENCED ASSESSMENT.....	8
1-4 PRINCIPLES OF ASSESSMENT.....	8
1-4-1 VALIDITY.....	8
1-4-2 RELIABILITY.....	9
1-4-3 PRACTICALITY.....	11
1-4-4 WASHBACK.....	11
1-4-5 AUTHENTICITY.....	11
1-5 APPROACHES TO ASSESSMENT: A BIT OF HISTORY.....	12
1-6 THE SPEAKING SKILL.....	15
1-7 ASSESSING SPEAKING.....	18
1-7-1 CHALLENGES AND DIFFICULTIES OF SPEAKING ASSESSMENT.....	19
1-7-2 TYPES OF SPEAKING TESTS.....	21
1-7-2-1 INDIRECT TESTS.....	21
1-7-2-2 DIRECT TESTS.....	21
1-7-2-3 SEMI-DIRECT TESTS.....	22

1-7-2-4 CONSTRUCT-BASED VERSUS TASK-BASED ASSESSMENT.....	23
1-7-2-5 STAND-ALONE VERSUS INTEGRATED ASSESSMENT.....	24
1-7-3 SPEAKING ASSESSMENT TASKS.....	24
1-7-4 SCORING SPEAKING TESTS.....	28
1-7-5 APPLYING PRINCIPLES TO SPEAKING TESTS.....	31
1-8 CONCLUSION.....	34

## **CHAPTER TWO: DATA COLLECTION**

<b>2-1 INTRODUCTION.....</b>	<b>36</b>
<b>2-2 THE ALGERIAN EDUCATIONAL SITUATION.....</b>	<b>36</b>
2-2-1 ENGLISH LANGUAGE TEACHING AT THE MIDDLE LEVEL.....	37
2-2-2 ENGLISH LANGUAGE TEACHING AT THE SECONDARY LEVEL.....	37
2-2-3 ENGLISH LANGUAGE TEACHING AT THE UNIVERSITY LEVEL.....	38
<b>2-3 RESEARCH METHODOLOGY.....</b>	<b>42</b>
2-3-1 QUANTITATIVE RESEARCH APPROACHES.....	42
2-3-1-1 DESCRIPTIVE SURVEY RESEARCH.....	43
2-3-1-2 EXPERIMENTAL RESEARCH.....	43
2-3-1-3 CAUSAL-COMPARATIVE RESEARCH.....	43
2-3-1-4 CORRELATIONAL RESEARCH.....	43
2-3-1-5 META-ANALYSIS RESEARCH.....	43
2-3-2 QUALITATIVE RESEARCH APPROACHES.....	43
2-3-2-1 ETHNOGRAPHIC STUDY.....	44
2-3-2-2 GROUNDED THEORY RESEARCH.....	44
2-3-2-3 PHENOMENOLOGY RESEARCH.....	44
2-3-2-4 CASE STUDY.....	44
<b>2-4 RESEARCH DESIGN.....</b>	<b>45</b>
2-4-1 RESEARCH OBJECTIVES.....	45
2-4-2 SAMPLING.....	46
2-4-3 DATA COLLECTION.....	48
2-4-3-1 THE OBSERVATION.....	49
2-4-3-2 THE STUDENTS' QUESTIONNAIRE.....	50
2-4-3-3 THE TEACHERS' INTERVIEW.....	51
<b>2-5 CONCLUSION.....</b>	<b>53</b>

## **CHAPTER THREE: DATA ANALYSIS AND INTERPRETATION**

<b>3.1 INTRODUCTION.....</b>	<b>54</b>
<b>3.2 DATA ANALYSIS.....</b>	<b>54</b>
3.2.1 THE QUESTIONNAIRE ANALYSIS.....	54
3.2.2 TEACHERS' INTERVIEW ANALYSIS.....	65
3.2.3 OBSERVATION ANALYSIS.....	69
<b>3.3 DATA INTERPRETATION.....</b>	<b>73</b>
<b>3.4 CONCLUSION.....</b>	<b>77</b>

## **CHAPTER FOUR: SUGGESTIONS AND RECOMMENDATIONS**

<b>4-1 INTRODUCTION.....</b>	<b>78</b>
<b>4-2 INTRODUCING SPEAKING AND SPEAKING ASSESSMENT.....</b>	<b>78</b>
4-2-1 PERFORMANCE-BASED ASSESSMENT.....	83
4-2-2 SELF AND PEER-ASSESSMENT.....	84
4-2-3 PORTFOLIO ASSESSMENT.....	85
<b>4-3 RATING CRITERIA.....</b>	<b>86</b>
<b>4-4 ENSURING A POSITIVE WASHBACK ASSESSMENT.....</b>	<b>88</b>
<b>4-5 SPEAKING ASSESSMENT TASKS.....</b>	<b>90</b>
<b>4-6 LESSENING THE DIFFICULTIES ENCOUNTERED BY THE STUDENTS.....</b>	<b>94</b>
4-6-1 STUDENTS' LACK OF VOCABULARY.....	94
4-6-2 THE PERSONALITY OF THE TEACHER.....	96
4-6-3 STUDENTS' ANXIETY.....	98
<b>4-7 CONCLUSION.....</b>	<b>100</b>

<b>GENERAL CONCLUSION .....</b>	<b>101</b>
---------------------------------	------------

<b>BIBLIOGRAPHY .....</b>	<b>104</b>
---------------------------	------------

<b>APPENDICES.....</b>	<b>113</b>
------------------------	------------

<b>APPENDIX 1.....</b>	<b>113</b>
------------------------	------------

<b>APPENDIX 2.....</b>	<b>115</b>
------------------------	------------

<b>APPENDIX 3.....</b>	<b>116</b>
------------------------	------------



## LIST OF ACRONYMS

<b>ACTFL</b>	: American Council for the Teaching of Foreign Languages
<b>AET</b>	: ARELS Examinations Trust of Oxford
<b>CBA</b>	: Competency-Based Approach
<b>CEFR</b>	: The Common European Framework of Reference for Languages
<b>CELS</b>	: The Cambridge Certificate in English Language Speaking
<b>CPE</b>	: Certificate of Proficiency in English oral interviews
<b>DRT</b>	: Directed Response Task
<b>EFL</b>	: English as a Foreign Language
<b>ELT</b>	: English Language Teaching
<b>ESP</b>	: English for Specific Purposes
<b>FCE</b>	: Cambridge First Certificate in English
<b>FSI</b>	: Foreign Service Institute
<b>IELTS</b>	: International English Language Testing System
<b>ILR</b>	: Interagency Language Roundtable
<b>LMD</b>	: License Master Doctorate
<b>OAS</b>	: Oral Assessment System
<b>OPI</b>	: Oral Proficiency Interview
<b>ROPE</b>	: Recorded Oral Proficiency Examination
<b>SOPI</b>	: Simulated Oral Proficiency Interview
<b>TEEP</b>	: Test in English for Educational Purposes
<b>TLU</b>	: Target Language Use
<b>TSE</b>	: Test of Spoken English
<b>UCLES</b>	: University of Cambridge Local Examination Syndicate

## LIST OF TABLES

<b>Table 2.1</b>	: The Programme of the First Four Semesters.....	39
<b>Table 2.2</b>	: The Programme of the Fifth and the Sixth Semesters.....	40
<b>Table 3.1</b>	: Reasons for Learning English.....	54
<b>Table 3.2</b>	: The Language Used by the Students to Speak With Their Colleagues.....	59
<b>Table 3.3</b>	: The Language Used by the Students to Speak With Their Teachers.....	59
<b>Table 3.4</b>	: Speaking With Native Speakers.....	60
<b>Table 3.5</b>	: Preferred Way of a Test Interview.....	63
<b>Table 3.6</b>	: Tasks Included In Teaching and Testing.....	66
<b>Table 3.7</b>	: Difficulties in Teaching and Testing Speaking.....	67
<b>Table 4.1</b>	: Common Reference Levels: Self-Assessment Grid.....	87

## LIST OF BAR GRAPHS

<b>Bar graph 3.1:</b> The Meaning of Learning a Language.....	55
<b>Bar graph 3.2:</b> The Skills Preferable to Master More by the Students.....	56
<b>Bar graph 3.3:</b> Preferred Tasks in the Classroom.....	58
<b>Bar graph 3.4:</b> Factors Affecting the Students' Speaking Performance.....	63
<b>Bar graph 3.5:</b> Criteria Used to Assess the Students' Speaking Ability.....	68
<b>Bar graph 3.6:</b> Test Tasks.....	72

## LIST OF PIE CHARTS

<b>Pie chart 3.1:</b> Students' Satisfaction with their Levels of Speaking.....	57
<b>Pie chart 3.2:</b> Students' Satisfaction with the Way of Teaching Speaking.....	58
<b>Pie chart 3.3:</b> Fluency vs. Accuracy.....	61
<b>Pie chart 3.4:</b> Attitudes towards Speaking Tests.....	61
<b>Pie chart 3.5:</b> Feelings during Speaking Tests.....	62
<b>Pie chart 3.6:</b> Students' Satisfaction with the Teachers' Ways of Rating.....	64
<b>Pie chart 3.7:</b> Types of Tests the Teachers Use.....	71

## **General Introduction**

In an era of a communication revolution in all fields over the world, English, beyond dispute, has become the global language that plays a crucial role in facilitating different treatments and formulating a bridge of communication between people. For this reason, English literacy becomes a matter of interest for the worldwide educational institutions, and the Algerian institutions are no exception. However, teaching and learning English in Algeria still encounter many obstacles that prevent learners from achieving high language proficiency.

Speaking in English is at the very heart of what it means to be a successful learner of this language. People, everywhere if they want to communicate with strangers, they, automatically, ask them, “*Do you speak English?*” they do not ask them if they can write, read, or listen to that language. Hence, speaking in English is of paramount importance and is, without doubt, the wish of every learner of the English language. This wish can become true through effective teaching and learning practices. However, we cannot judge this efficiency except by following good assessment procedures. Since speaking is considered a complicated skill that involves many sub-skills, assessing speaking is complicated too. The present study aims at investigating the assessment of second year license students’ speaking skill at the University of Tlemcen through an attempt to answer the following questions:

- 1- What do teachers assess in a speaking performance?
- 2- How do they assess their students’ ability to speak?
- 3- What are the linguistic and psychological difficulties that hinder the students’ speaking abilities?

In order to answer these questions, the researcher proposed the following hypotheses:

- 1- Teachers assess the speaking as a productive skill, not as an interactive skill
- 2- Teachers are still using the out of date tasks and rating criteria in testing the students’ speaking ability.

- 3- The linguistic and psychological difficulties that hinder the students' speaking ability are anxiety and lack of vocabulary

This research work attempts through four chapters to redress the issue of speaking assessment and find out the possible remedies to the difficulties found either by the teachers or by the students when doing such an assessment. Chapter one highlights the literature review of the concept “*assessment*” in general and “*speaking assessment*” in particular. It provides definitions that enable the reader of distinguishing between three confusing terms *assessment*, *testing*, and *evaluation*. Then, explains in some details the different principles of a good assessment: reliability, validity, practicality, authenticity, and washback. After that moves to the assessment of the speaking skill, mentioning the definition of “speaking” from different perspectives, the types of speaking assessments, the challenges that face test designers when designing a speaking test, and the principles of a speaking assessment.

Chapter two revolves around the empirical side of the current study. First, it throws the light on the English teaching situation in the middle, secondary and the tertiary levels in the Algerian institutions. Then, it provides an overview on the research methodology followed by the researcher, which is a case study combining both quantitative and qualitative data. After that, it provides a detailed description of the design of this research work starting with the research main objectives, then moving to the sampling description, and the instruments utilized to collect data for this study including the students' questionnaire, the teachers' interview and the test observation.

Chapter three analyses the collected data and interprets them. The researcher, when analysing the results, resorts to the use of tables and graphics such as pie charts and diagrams to illustrate the quantitative data, and uses expressions such as ‘few’, ‘the majority’ when qualitative data are concerned. The interpretation of the results leads the researcher to come across the main findings, which will answer the aforementioned questions and will either strengthen or weaken the hypotheses put previously by the researcher.

Chapter four, in light of the findings provided in the preceding chapter, proposes some suggestions and recommendations that help alleviate the problems. It provides

some definitions for the term ‘speaking’ that present a basis for selecting the relevant criteria to assess the students’ speaking ability. It also suggests the test type the more appropriate for assessing speaking objectively, the tasks that are likely to motivate the students to show their ability of speaking as well as some cures to the difficulties that students encounter either in learning or when being assessed.

# CHAPTER ONE

## ASSESSING SPEAKING A LITERATURE REVIEW



## **ASSESSING SPEAKING: A LITERATURE REVIEW**

<b>1-1 INTRODUCTION.....</b>	<b>4</b>
<b>1-2 DEFINITION OF BASIC CONCEPTS: Evaluation, Assessment, And Testing.....</b>	<b>4</b>
<b>1-3 TYPES OF ASSESSMENT.....</b>	<b>5</b>
1-3-1 SUMMATIVE AND FORMATIVE ASSESSMENT.....	5
1-3-2 FORMAL AND INFORMAL ASSESSMENT.....	6
1-3-2-1 FORMAL ASSESSMENT.....	6
1-3-2-2 INFORMAL ASSESSMENT.....	7
1-3-3 NORM-REFERENCED AND CRITERION-REFERENCED ASSESSMENT.....	8
<b>1-4 PRINCIPLES OF ASSESSMENT.....</b>	<b>8</b>
1-4-1 VALIDITY.....	8
1-4-2 RELIABILITY.....	9
1-4-3 PRACTICALITY.....	11
1-4-4 WASHBACK.....	11
1-4-5 AUTHENTICITY.....	11
<b>1-5 APPROACHES TO ASSESSMENT: A BIT OF HISTORY.....</b>	<b>12</b>
<b>1-6 THE SPEAKING SKILL.....</b>	<b>15</b>
<b>1-7 ASSESSING SPEAKING.....</b>	<b>18</b>
1-7-1 CHALLENGES AND DIFFICULTIES OF SPEAKING ASSESSMENT.....	19
1-7-2 TYPES OF SPEAKING TESTS.....	21
1-7-2-1 INDIRECT TESTS.....	21
1-7-2-2 DIRECT TESTS.....	21
1-7-2-3 SEMI-DIRECT TESTS.....	22
1-7-2-4 CONSTRUCT-BASED VERSUS TASK-BASED ASSESSMENT.....	23
1-7-2-5 STAND-ALONE VERSUS INTEGRATED ASSESSMENT.....	24
1-7-3 SPEAKING ASSESSMENT TASKS.....	24
1-7-4 SCORING SPEAKING TESTS.....	28
1-7-5 APPLYING PRINCIPLES TO SPEAKING TESTS.....	31
<b>1-8 CONCLUSION.....</b>	<b>34</b>

## CHAPTER ONE

### Assessing Speaking: A Literature Review

#### 1-1 Introduction

Chapter one sheds the light on the theoretical side of this research work. It is divided into two main parts: the first part will be concerned with assessments in general and the second part with the assessment of the speaking skill in particular. The first part first introduces the definition of the key concepts of the study. Then, it mentions in some detailed way the main principles of tests and testing. After that, it provides a historical point of view of testing and assessments. The second part firstly investigates the nature of the term speaking and provides different viewpoints. Then, it deals with the assessment of this skill mentioning the types of speaking tests, the challenges that face test developers in designing such tests, the different tasks often used to elicit the desired performances, as well as the principles underlying the speaking tests. Within the principles, it discusses about the reliability of a speaking test and the factors that may influence it. It also introduces the validity of speaking tests, namely the construct validity because it is the most important for this type of tests. Lastly, it talks briefly about practicality and washback of speaking tests.

#### 1-2 Definition of Basic Concepts: Evaluation, Assessment, and Testing

In the educational field, assessment and testing are usually used interchangeably. They are used to mean measuring the progress of learners. However, assessment is broader than testing. *Assessment* is a continuing process of informing learners about their achievements using several ways, while testing is the formal way to do this (Pawlak & Waniek-Klimczak, 2015). “**Testing** is not, of course, the only way in which information about people’s language ability can be gathered. It is just one form of assessment, and other methods will often be appropriate.” (Hughes, 2003, p. 5). Carroll (1968) defines a test as “a procedure designed to elicit certain behaviour from which one can make inferences about certain characteristics of an individual” (As cited in Bachman, 1990, p. 20). Testing is the only way to give the same opportunity to all learners to show their abilities. Testing is the means by which we guarantee *fairness* (Douglas, 2010).

*Evaluation*, on the other hand, is also an ongoing process of collecting data, not for giving feedback to learners as assessment, but for making decisions (Weiss, as cited in Bachman, 1990). Making the right decision depends on the reliability and the relevance of the collected information. This information can be either quantitative such as test scores or qualitative such as observations, letters of reference, or overall impressions (Bachman, 1990). Tests are evaluative only when it comes to ranking learners basing on their results. Otherwise, they have pedagogical purposes such as motivating learners or giving feedback. In other words, “not all tests are evaluative, and not all evaluation involves tests” (Bachman, 1990, p. 24). Douglas claims that tests “can help confirm our own assessments and help us make decisions about students’ needs with more confidence” (2010, p. 1). He joins and mentions the relationship between the three concepts: test, assessment and evaluation.

### **1-3 Types of Assessment**

Researchers distinguish between different ways of assessment. This distinction may depend on the information sought behind a given assessment, the way assessment is performed, and the purpose or the function of the assessment. In this sense, the distinctions are mentioned:

#### **1-3-1 Summative and Formative Assessment**

A distinction is made between summative and formative assessment. Summative assessment or “*assessment of learning*” is a graded assessment carried out at the end of a course or programme of study in order to measure what the learners have learnt “... for purposes of certification and (latterly) accountability” (Torrance & Pryor, 1998, p. 8). Final and proficiency exams are examples of summative assessment. However, Formative Assessment or “*assessment for learning*” is often a non-graded assessment that takes place during a course aiming at promoting the level of students by providing them with useful feedback,

A key development in the formative assessment process, and an indispensable condition for improvement, is that the student comes to hold a notion of the standard or desired quality similar to that of the teacher, is able to monitor the quality of what is being produced at the time of production, and is able to regulate their work appropriately (Gipps, 1994, p. 126)

The positive point is the fact that learners can adjust their learning, before it is too late.

### 1-3-2 Formal and Informal Assessment

Assessments could be performed either in a formal way, that is what we call *tests*, or in an informal way, using different methods.

#### 1-3-2-1 Formal Assessment

Formal assessment is performed under special classroom conditions. It is evaluated by giving grades to each student's work. There are four main types of test: Placement test, Achievement test, Proficiency test, and Diagnostic test.

A *placement test* is needed to place learners in their right classes according to their levels "The aim is to produce groups which are homogeneous in level that will use institutional and teacher time most effectively" (Harris & McCann, 1994, p. 29). It has to be noted in this type of assessment that a placement test should be performed at the beginning of the year or the course as its function is to put the learners in their right classes. *An achievement test* is performed at the end of a course to determine how much learners have learnt, or, as its name indicates, what learners have achieved at the end of a course, "...their purpose being to establish how successful individual student, groups of students, or the courses themselves have been in achieving objectives" (Hughes, 2003, p. 13).

There are two kinds of achievement tests: *Final achievement and Progress achievement test*. *The final achievement test* is performed at the end of a study course. This kind of tests may be constructed by the education ministry, authoritative examining board, or by members of instructional establishments. However, the *progress achievement test* is intended to find out the progress made by learners and take the right decision on what to be taught next. It is administered during the courses (Hughes, 2003). Progress tests "can provide a great deal of information if the test samples widely from the course content" (Harris & McCann, 1994, p. 28). A *diagnostic test* is the type of tests that determine which area of language creates difficulties for the learners to "... design further course activities and work out remedial activities" (Harris & McCann, 1994, p. 29). Investigating these kinds of tests, one would make no distinction between diagnostic and placement test. Alderson argues that,

... However, diagnostic tests are frequently confused with placement tests...

Both diagnostic and placement tests appear to be designed to identify what a learner knows in order to decide on future teaching or learning. Placement tests,

after all, are intended to group learners in homogeneous groups in order to have a suitable basis for further teaching and learning (2005, p. 4)

The main difference between the two types is that a placement test is used to place students according to their level, thus it can be considered as a diagnostic test while the diagnostic test has to be performed many times during the course.

**Proficiency Test**, on the other hand, is not related to the learning subjects. It attempts to measure a person's general ability, competence, or mastery of a foreign language without taking into account any course that a person may have taken before, "It is based on a specification of what candidates have to be able to do in the language in order to be proficient." (Hughes, 2003, p. 11). In fact, a proficiency test is considered as summative and criterion-referenced test (Hughes, 2003).

### 1-3-2-2 Informal Assessment

**Informal assessment** is done through observing the learners' performance in the classroom circumstances. It is often referred to as *alternatives in assessment* or *continuous assessment (CA)*. Informal assessment gives feedback to the students to support them do better in their tests or *formal assessment* (Puhl, 1997). Furthermore, informal assessment not only helps students to better their test results, but also

Holds rich potential for teachers because it affirms high-order creative and critical thinking and because it embraces not only cognitive outcomes but affective and behavioural outcomes as well.....CA in practice can embody the global changes that effect the very nature of the classroom process, bringing it away from education as information and toward the full development of learner potential (Puhl, 1997, p. 3)

One can conclude that informal assessments are useful even for the long-life learning. Informal assessment may be performed in different ways. Some examples of informal assessment are Self-assessment, peer- assessment, and portfolio-assessment. *Self* and *peer-assessment* enhance self-control over learning and prompt the learners to understand, and be

involved in, their learning. Yet, teachers have to work hard to get their learners understand and generate these devices (Puhl, 1997).

A *portfolio* is a set of student activities collected all over a period of time. Self and peer-assessment are part of the portfolio assessment. “It is generally recognized that one of the main benefits of portfolio assessment is the promotion of learner reflection” (Santos, 1997, p. 10). Reflecting means that learners sift through their previous classroom works and note their weaknesses and strengths in order to improve their learning. This reflection helps playing a part in the triangulation of evidence gathered for the assessment procedure (Santos, 1997)

### **1-3-3 Norm-referenced and Criterion-referenced Assessment**

In norm-referenced tests, a learner’s language ability is compared with that of his or her fellows. For example, when a student performs a speaking norm-referenced test, he/she will obtain “a score that place him/her in the top 10 per cent of candidates who have taken that test... we are not told directly what the student is capable of doing in the language” (Hughes, 2003, p. 20). The beneficial use of norm-referenced tests is in placement tests. However, in a criterion-referenced test, the candidates’ performance is assessed in relation to some criteria put by the scorer or by the educational institutions then the candidates who perform well pass (Hughes, 2003). Hughes adds that “Criterion-referenced tests have two positive virtues: they are meaningful standards in terms of what people can do, which do not change with groups of candidates, and they motivate students to attain those standards” (2003, p. 21). Criterion-referenced tests, as stated by Hughes, are beneficial in conveying the real language ability of the learners, which is constant and does not alter among different groups.

### **1-4 Principles of Assessment**

There are five necessary principles for any test to be fair and useful; validity, reliability, practicality, washback, and authenticity.

#### **1-4-1 Validity**

A test is valid when it assesses the skill or the language area that is supposed to be assessed; Fulcher and Davidson explain that,

This view of validity presupposes that when we write a test we have an intention to measure something, that the ‘something’ is ‘real’, and that validity enquiry concerns finding out whether a test ‘actually does measure’ what is intended. (2007, p. 4)

Though there is no precise method, the validity of a test can be recognised using different types of evidence. *Content validity* is one of the evidence that helps establishing the validity of a test. It is the degree of matching between the test content and the test objectives. *Criterion-related validity* is another evidence that is related to the degree of agreement between the results obtained and some criterion measures (Brown, 2004). *Face validity* is also a useful sign. It is linked to the appearance of the test. It is “the degree to which a test looks right, and appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of the examinees who take it” (Brown, 2004, p. 26). Another confirmation of a test validity is *construct validity*, which means the interpretation of scores given by the scorer as a proof of the testees’ language ability. Bachman and Palmer state that;

.... We need to demonstrate, or justify the validity of scores and not simply assert or argue that they are valid... The term construct validity is therefore used to refer to the extent to which we can interpret a given test score as an indicator of the ability (ies) or construct(s), we want to measure (1996, p. 21)

The term ‘*construct*’ means any characteristic, ability or skill that occurs in the human brain and is proved by practice. For instance, speaking is a construct that is shown through practising speaking (Brown, 2004).

Validity is not the only principle that characterizes tests. For a test to be valid, it is necessary to “provide accurate measurements. It must therefore be reliable” (Hughes, 2003, p. 50). Thus, reliability is another principle of assessment.

### 1-4-2 Reliability

A reliable test shows similar testees’ scores if it is repeated under the same conditions. In other words, if a test is performed today by a group of students, then performed again tomorrow by the same group, the scores of both performances should be very similar to each other, otherwise there must be a problem of reliability. Reliability is related to the degree of

representativeness of results (Brown, 2004). Chapelle suggests that “The classic role of reliability is as but one form of evidence of the validity of a test; this is congruent with older arguments that a test must be reliable before it is valid. ... Reliability becomes an element in a validity argument” (1994, as cited in Davidson & K. Lynch, 2002, p. 134).

Brown (2004) points out that reliability can be best dealt with by addressing some factors that have a negative impact on it as they lead to the unreliability of the test. This impact is represented by the fluctuations in the student himself (or what is called student-related reliability), in the scoring (or what is called rater-reliability), in the test administration (or what is called test-administration reliability), and in the test itself (or what is called test-reliability).

**Student-Related Reliability** is embodied in the anxiety and bad feeling that accompany students during their days of testing. This affects negatively the real ability of the students. Likewise, some temporary psychological or physical illnesses may have their bad influences. mentions an example of this problem,

The learner’s motivation is also a very important factor that deeply affects reliability since high and continuous motivation will cause the test to be reliable while low motivation causes a low reliability. But this motivation cannot be controlled since.... some of the learners may have a low motivation because of some psychological or personal problems (Zidane, 2010, p. 25)

As such, student-related reliability is uncontrolled since it is related to humans’ attitudes. However, there must be some remedies to alleviate such a problem.

**Rater Reliability** is related to the scorer who is likely to commit some errors in scoring, in addition to subjectivity, inexperience, and inattention. There are two kinds of scorer’s reliability: Inter-rater and intra-rater reliability. *Inter-rater reliability* occurs when two or more scorers show similar results of the same test while *intra-rater reliability* or *test-retest reliability* arises when the test is performed twice by the same testees, scored by the same scorer, and the obtained results are alike. Unreliability in such a case may be the result of fatigue, bias, or carelessness (Brown, 2004).

**Test-Administration Reliability** refers to the conditions under which the test is administered and that may affect test reliability such as noise, bad quality of photocopies, bad light, irrelevant temperature, and uncomfortable desks and chairs (Brown, 2004). However,



**Test Reliability** has a relation with the factors that may generate errors in measurement such as the content, the time limit and the nature of the test itself (Brown, 2004).

### **1-4-3 Practicality**

A practical test is a test, which is easy to administer, to score and to interpret. A practical test should not be costly in terms of time and material resources, including money (Brown, 2004). Bachman and Palmer affirm that:

.. Determining the practicality of a given test involves the consideration of (1) the resources that will be required to develop an operational test that has a balance of qualities we want, and (2) the allocation and management of the resources that are available (1996, p. 36)

Accordingly, any test requiring high costs, much time to perform or to evaluate, is considered as impractical.

### **1-4-4 Washback**

It is the impact, either positive or negative that tests have on learners, teachers, educational systems, and societies. This impact is often seen as being negative (Hughes, 2003). Benmostefa adds that a washback is “an inherent quality of any kind of assessment, especially when pupils and students’ futures are affected by the examination results.... When learners hear the word TESTING, they feel nervous and tense” (2014, p. 166). High-stakes tests, for example, may affect negatively the curriculum in that teachers consume much time teaching to the test items rather than to the skill that should be tested (Gipps, 1994).

### **1-4-5 Authenticity**

It is the degree to which test items reflect real life. In an authentic test, the language is natural and topics are extracted from real life to generate a motivation that enables the test takers to get better grades (Brown, 2004). The criterion of authenticity is indispensable in designing test items because it “raises important pragmatic and ethical questions in language testing...Lack of authenticity in the material used in a test raises issues about the generalizability of results” Spolsky (1985, as cited in Bachman, 1990, p. 301). In other words, authenticity is

the extent of resemblance between test task and the target language use, and thus the scores obtained in a test should reflect not only the examinee's performance during the test, but also the examinee's use of the language outside the test (Bachman, 1996).

## 1-5 Approaches to Assessment: An Overview of History

Language testing and assessment have been going along with the development that occurred in language teaching or the shifting sands of language teaching methodology (Brown, 2004). Benmostefa (2014) affirms that language testing have witnessed an increasing maturity throughout the last few decades. Four main approaches, or periods dominated the language-testing field. In fact, there was no time split between those periods; they were overlapping, but many scientists listed them one after the other.

The first period of testing was *The Pre-scientific period*. It dominated the first half of the 20<sup>th</sup> century. It was characterized by the principles underlying the Grammar-Translation method. Hence, the adopted approach was *The Essay-Translation Approach*. The main activities within this approach were translation from the target language into mother tongue and vice versa, grammar tasks, and essay writing (Benmostefa, 2014).

The second period played a part in the early 1950s and lasted about two decades. It is *The Psychometric-Structuralist Period* when tests were for the first time designed on a premise of scientific trends. It was stamped by the principles of structuralism and behaviourism appearing at that time. The approach adopted was *The Discrete-Point Approach* (Alduais, 2012). It was based on the assumption that a language can be divided into its small components: the four skills, phonology, morphology, lexicon and syntax. Each skill can be tested in isolation. According to Brown, this approach to testing focuses “on specific language elements such as the phonological, grammatical, and lexical contrasts between two languages” (2004, p. 8). It has a trend to *decontextualize* the knowledge to be tested (Alduais, 2012).

One of the main advantages of the discrete-point testing is its high objectivity resulting in an unquestionable test reliability. However, the main disadvantage is its decontextualization that made it far from achieving a natural verbal communication (Alduais, 2012). Brown argues that:

Language competence is a unified set of interacting abilities that cannot be tested separately.....communicative competence is so global and requires such integration (hence the term "integrative-testing") that it cannot be captured in

additive tests of grammar, reading, vocabulary and other discrete-points of language (2004, p. 8)

These trends influenced language teaching and testing. Consequently, a new approach had been brought about. It was *The Integrative-Sociolinguistic Approach*, also called the *Psycholinguistic-Sociolinguistic Approach*. The main theoretical premise behind this approach is that language cannot be divided into its smallest components; it must be tested as a one unit. Oral interview, writing composition, dictation, and cloze tests are examples of integrative testing (Alduais, 2012).

Once again, the integrative approach to testing was criticized “because no clear-short-cuts were drawn regarding testing communicative skills of language” (Alduais, 2012, p. 205). Weir (1990) claims that tasks within that approach were unspontaneous and far from reflecting real life language. He adds, “Integrative tests such as cloze tests only tell us about a candidate's linguistic competence. They do not tell us anything directly about a student's performance ability” (As cited in Brown, 2004, p. 10).

Hymes (1972) introduced *the communicative competence*, which he defined as being the capability to use linguistic competencies in different communicative contexts. Thus, he dealt with both *usage* and *use*. Hymes's theory was the linguistic background to **The Communicative Language Testing Approach**, which involves knowledge and performance of the language. It considers that grammatical forms are learnt to reach a meaningful communicative performance (Bagarić & Djigunović, 2007, p. 94). The distinction that can be made between the psychometric-structuralists, integrative and communicative approaches is that the two first approaches consider language as a psychological fact, while the third one interprets it as a social fact (Alduais, 2012).

Many attempts were offered to construct communicative competence tests. Canale and Swain model, Bachman and Palmer model and the Common European Framework (CEF) are some examples of these attempts. Fulcher and Davidson agree that:

All models of language ability have three dimensions, constituted by statements about... what it means to know a language (a model of knowledge)... underlying factors relating to the ability to use language (a model of performance)... how we understand specific instances of language use (actual language use) (2007, p. 37)

The communicative competence in **Canale and Swain model (1980)** refers ‘*to the interaction between grammatical competence, or knowledge of the rules of grammar, and sociolinguistic competence, or knowledge of the rules of language use*’ (Canal and Swain, 1980 as cited in Fulcher and Davidson, 2007, p. 37). This communicative competence consists of *grammatical competence* (grammar, lexis, morphology, syntax, semantics, and phonology), *sociolinguistic competence* (knowledge of the sociolinguistic rules), and *strategic competence* (knowledge of the strategies that help the learner to use alternative utterances in communication).

Due to its simplicity, The Canal and Swain model lasted and dominated for more than a decade, “The easiness with which the model of Canale and Swain can be applied is probably the main reason why many researchers of communicative competence still use it” (Bagarić & Djigunović, 2007, p. 87). Nevertheless, a new model; **the Bachman and Palmer Model (1996)** came to appear. In 1990, Bachman suggested a new model of communicative competence. This model was developed in the mid 1990’s. It was based on two main elements: *language knowledge* and *strategic competence*. Bachman and Palmer mean by *language knowledge* the amount of information about language that is stored in the human memory to be used when needed, and by *strategic competence* the skill of relating the language knowledge to the context of the language use and to the background of the language user (Bachman and Palmer, 1996). Moreover, they assert that there are,

Other individual characteristics that we also need to consider are personal characteristics, topical knowledge, and affective schemata. We include these in our discussion for two reasons. First, we believe that these characteristics can have important influence on both language use and test performance. Second, we believe that it is possible and desirable to design language tests so that these characteristics facilitate rather than impede test takers’ performance (Bachman & Palmer, 1996, p. 61)

Bachman and Palmer mean by *personal characteristics* all the features that distinguish a person from another such as age, sex, nationality, native language, resident status, level and type of general education, and prior experience with a given test (Bachman and Palmer, 1996).

As for topical knowledge “We might expect students of economics or finance to bring relevant topical knowledge to a writing test that request them to discuss the advantages and disadvantages of a liberal fiscal policy” (Bachman and Palmer, 1996, p. 61)

According to Fulcher and Davidson (2007), *affective schemata* are the previous experiences of a test taker that yield some emotions towards a given test topic, so that he/she deals with it with a high or a low motivation. To illustrate this concept, Bachman and Palmer (1996) introduced some examples of test tasks that ask the test takers to discuss topics such as gun control or national sovereignty. The testees’ emotional reaction towards these topics affect their performance. That is to say, the positive feelings towards a favoured topic would affect positively the test takers’ responses, as they are enthusiastic about it.

### 1-6 The Speaking Skill

Speaking is the most important skill that is used to facilitate communication between individuals. Speaking is the communicative vehicle by which we express ourselves, discuss with each other, learn, and make relationships (Bygate, 1987). In educational contexts, “...acquiring speaking proficiency is one of the hardest skills for ELL... to achieve” (Sasson, 2013, p. 24). Lado asserts that “To be able to speak a foreign language is, without doubt, the most highly prized language skill” (1961, as cited in Fultcher, 2003, p. 18). Khamkhien (2010) deems that among the four known language skills, speaking is the most important to be learnt. The speaking skill encompasses all the other language skills (Ur, 1996, as cited in Khamkhien, 2010). Bailey mentions that speaking is the most important skill to learn because “people who know a language are referred to as a “speaker” of the language as though speaking included all other skills” (2005, as cited in, Tuttle & Tuttle, 2013, p. 3). Luoma affirms that:

The ability to speak in a foreign language is at the very heart of what it means to be able to use a foreign language. Our personality, our self-image, our knowledge of the world and our ability to reason and express our thoughts are all reflected in our spoken performance in a foreign language (2004, p. ix)

People often ‘*speak*’ to express themselves, interact with others, make relationships, defend themselves and their opinions, learn, travel, buy and sell, and to meet their needs.

However, the speaking skill is not given importance due to its momentary and spontaneous nature (Bygate, 1987). Most learners, owing to the little importance given to teaching speaking, lost confidence and motivation because they are unable to speak fluently or to express themselves. A common phenomenon is spread among language learners is that many students who are excellent in the other subject matters, can hardly speak the foreign language (Tuttle & Tuttle, 2013).

From a scientific standpoint, a skilful speaker is a speaker who masters two sub-skills: motor-perceptive skills and interaction skills. **Motor-Perceptive Skills** refer to the different processes that take place in the human mind to produce a right speech. These processes include the perception, the recall, and the articulation of sounds employing grammar, vocabulary, and phonology. **Interaction skills** are the skills used to speak effectively when taking part in an interaction. They include decisions about what to say and how, and also how to avoid interactional troubles (Bygate, 1987).

Luoma (2004, p. 9) talks about four features that underlie speaking: *pronunciation, spoken grammar, spoken vocabulary, and the sociolinguistic application of speech*. **Pronunciation** is the sound of speech. Though it is obvious that the native speaker is the standard of pronunciation, the effectiveness of communication, which is based on comprehensibility, is a better standard for learner pronunciation. **Spoken Grammar** is different from written grammar. Writers use sentences; speakers use idea units which are short clauses or phrases. Idea units are sometimes used without verbs, or are begun by the speaker and completed by the listener. Consequently, the spoken grammar is simple and does not require much attention as is the case when writing, except in specific cases where speech is prearranged such as lectures or conferences.

Hilliard (2014, p. 3) lists six features underlying the spoken grammar *ellipsis, heads, tails, fillers, backchannels, and phrasal chunks*. **Ellipsis** is the omission of some items e.g. “any question?” Instead of “Do you have any question?” **Heads**, also called **left-dislocation**, is to talk about the main topic before giving any information to draw the attention of the listener, e.g. “The party last night, it was fantastic” instead of “The party last night was fantastic”. **Tails**, also called **right-dislocation**, are comments added at the end of the utterance, e.g. “she was very intelligent, the student from Algiers” (with tail), “The student from Algiers was very intelligent,” (no tail). **Fillers** such as *hmm, err, um, well* are words with no intended meaning serve to give the opportunity to the speaker to organize his or her ideas. **Backchannels** are

words such as *yeah, I see, oh* that serve to show that the listener is concentrating and to encourage the speaker to continue speaking. **Phrasal chunks** are fixed phrases serving at alleviating the pressure of time when speaking. *Sort of, kind of, a bit, a little bit, you know, and I mean* are some examples of phrasal chunks.

**Spoken vocabularies** have to be well chosen to make the message clear. The use of simple words is recommended for a meaningful interaction, as they are indicators of highly advanced speaking (Read, 2005, as cited in Luoma, 2004). Speakers must have a knowledge of “words, phrases and strategies for creating time to speak. These are sometimes called fillers or hesitation markers” (Luoma, 2004, p. 18). These involve words such as *ah, you see, you know*, and expressions such as *that’s a good idea, let’s think about, let me see*. The advantage in using these fillers is to get the opportunity and the time to think about what to say. Words in English are *specific* or *generic*. In a written passage, the writer uses **specific words** to enable the reader of understanding what is written, while in a conversational situation, speakers use **generic words** such as *this thing, that boy, go, come, well*. They are common among the speakers and make the communication fast and natural (Luoma, 2004).

From a **sociolinguistic** standpoint, speech varies from a situation to another. Situations are summarized in the framework **SPEAKING** proposed by Hymes. This framework is an acronym for: **Situation, Participants, Ends, Act, Key, Instrumentalities, Norms, and Genre**. **Situation** refers to where speaking occurs. **Participants** refer to the people sharing the speaking act. **Ends** mean the goals of the speech. **Act sequence** implies the content of the speech and its sequencing. **Key** signifies the atmosphere of the speaking act whether it is for example friendly or formal. **Instrumentalities** or channels of speech, which can be *spoken lively* or *recorded*. **Norms** of interaction, and **Genre** of interaction such as joke, story. (Luoma, 2004).

Thornbury (2005) makes a distinction between what speakers **do** and what they **know**. What speakers **do** refers to the speaking mental process, or how speech is produced. This process passes through three stages: *conceptualization, formulation and articulation*. *Conceptualization* occurs when concepts are prepared at the level of the mind to convey a message. *Formulation* refers to the stage when those concepts are organized in a linguistic form, including grammatical, morphological, and phonological encoding. *Articulation* is the executive stage in which organs of speech production such as lungs, teeth, lips are used to convey a meaningful message.

On the other hand, what speakers *know* refers to the *linguistic knowledge*, *extralinguistic knowledge*, and *sociocultural knowledge*. *Linguistic knowledge* comprises language features such as grammar, vocabulary and phonology. *Extralinguistic knowledge* encompasses cultural knowledge, knowledge of the context, and familiarity with the interlocutor. Sociolinguistic knowledge means to be aware of the social conventions (Thornbury, 2005).

### 1-7 Assessing Speaking

The dominance of the communicative approach to language testing has made an emphasis on the speaking skill and its assessment (Sak, 2008). The ability to speak a foreign language is difficult, and the ability to assess this speaking is challenging (Luoma, 2004). The speaking skill is the most troublesome skill to be assessed because it encompasses a mixture of many sub-skills that puzzle test designers over which method is better to ensure fair, reliable, valid, and practical tests (Kitao & Kitao, 1996, as cited in Trejos et al, 2013).

The assessment of speaking skills is classified as less important than the assessment of the other skills (Knight, 1992). Hughes (2003) states that at the beginning of a speaking course, it is not necessary to assess students formally, however, giving a feedback such as informal observation or information is rather more preferable. Lado (1961, p. 239) wrote:

The ability to speak a foreign language is without doubt the most highly prized language skill, and rightly so. . . . Yet testing the ability to speak a foreign language is perhaps the least developed and the least practiced in the language-testing field (as cited in Fulcher, 2003, p. 18)

Lado emphasises the necessity of providing a clear definition of speaking and developing the field of assessing the ability of speaking

The difficulty of assessing speaking stems from the nature of this skill as being a momentary procedure and time-consuming since examinees are assessed one after one. Moreover, the assessment of speaking is often subjective and affected by many aspects. This problem can be solved by implementing relevant rating scales (Pawlak & Waniek-Klimczak, 2015).



**1-7-1 Challenges and Difficulties of Speaking Assessment**

Regarding the complex nature of the speaking skill, test makers face many difficulties in designing effective speaking tests. Luoma asserts that the assessment of speaking is a thought-provoking issue because “there are so many factors that influence our impression of how well someone can speak a language, and because we expect test scores to be accurate, just and appropriate for our purpose” (2004, p. 1).

Alderson and Bachman state that ensuring reliability is most difficult when assessing speaking. The face-to-face interaction between the examinee and the interviewer requires the rater to judge the performance in real time. This means that the rater may forget many features and builds his or her judgement only on what remains in his or her memory such as pronunciation, accuracy and fluency at the cost of other important factors such as the language level, the relation between the interlocutor and the examinee, and personal characteristics (Alderson and Bachman’s preface to Luoma, 2004).

Knight raises the most challenging question among speaking testers: “What should we look for when we assess a student’s ability to speak English?” (1992, p. 294). The workshop deals with two issues that are essential to produce valid and reliable tests: the relevant criteria that should be looked for when testing learners’ speaking abilities, and the correlation between the selected criteria, their weighting and the context of the test. These criteria have to be explicit in order to increase the reliability of the assessor. Besides, He speaks about the problem of selecting the appropriate tasks that help eliciting the appropriate responses, and that of designing reliable tests, as well as that of save time and facilities when administering speaking tests. Hughes (2003, p. 8) lists three main problems that face testers. These are:

- 1- The selection of tasks that represent real life situations and meet the needs of the examinees.
- 2- The selected tasks should highlight the real ability of examinees.
- 3- The scoring should be valid and reliable.

Bachman (1990) mentions that the changes occurring in language teaching and learning have changed the views on language testing. This has raised the challenge to language test designers in terms of how can we know the factors that affect the students’ speaking performances through the use of tests, and how can we benefit from this knowledge to administer more effective tests.

From another perspective, Fulcher and Davidson (2007, p. 258) draw our attention to three kinds of speaking test problems. Firstly, “How to simulate real-world tasks in a test?” Secondly, can we obtain scores that are generalizable to real-world performance? In another word “to which extent test scores reflect the real speaking ability of the students outside the test, i.e. in real-life situations? Thirdly, in terms of standardization, and giving that a learner’s performance and interaction fluctuate with different interlocutors, how to proceed to provide the same test experience for all the examinees. Is it better to rely on tape-based tests as being more standardized or to provide a training for the interlocutors so that they interact with learners fairly? “A preference for a tape-mediated test is expressed by those who see variation by the interlocutor as a potential threat to validity” (Fulcher and Davidson, 2007, p. 258).

Luoma (2004) brings the same problem of interlocutors to light. She says that the personality and the communication style of the interviewer effect the performance of the examinee. She also mentions another problem related to the scoring procedure. Raters do not have an agreement about how a performance should be scored. This issue leads to a problem of reliability. However, it can be worked out by preparing a precise rating scale that would help raters to evaluate exactly the learners’ performances.

Brown (2004) points out that the assessment of the speaking skill faces several challenges. First, speaking is connected firmly with listening, especially in interactions where it is necessary to have an accurate and effective listening to produce speaking. It is difficult for the assessor to score a speaking performance without ensuring that the speaker has understood the message of the interlocutor. The second challenge is how to design tasks that can elicit the response which serve the purpose of the test, for example, if the examiner wants to assess the speaker’s grammar, the tasks included in the test must elicit the grammatical knowledge of the examinee. The third challenge is the difficulty of scoring an open-ended speaking.

From a psychological point of view, test anxiety is one of the most factors that influence learners’ performances in speaking exams. Researchers demonstrate that there exists a strong relationship between language anxiety and speaking tests. The nature of the speaking, as being a real-time skill, provokes examinees’ anxiety. It is the fact that speaking tests will determine a decisive grade that exercises pressure and nervousness on the examinees (Herwitt and Stephenson, 2011, as cited in Pawlak & Waniek-Klimczak, 2015). Kitano reveals that:

The majority of students believed speaking skills to be the most important; hence, they experienced more self-imposed pressure on being successful in this

area, which may have generated an additional amount of fear and tension during speaking activities (2001, as cited in Pawlak & Waniek-Klimczak, 2015, p. 106)

Kitano (2001, as cited in Pawlak & Waniek-Klimczak, 2015) points out that learners get pressure when they compare their own speaking with that of their peers. Dörnyei and Kormos (2000, as cited in Pawlak & Waniek-Klimczak, 2015) state that anxiety is contagious, i.e. the anxious examinee influences negatively the other examinees, and this affects their performance. The factors leading to language anxiety are lack of fluency, vocabulary, pronunciation and content (Pawlak & Waniek-Klimczak, 2015).

From a practical standpoint, Sak (2008) poses the problem of administration costs, the hardness of testing a large number of examinees, the responsibility of training the examiners, and the big amount of time consumed to test all the examinees.

### **1-7-2 Types of Speaking Tests**

Assessing the speaking skill of learners can be performed in three formats: *indirect*, *semi-direct* and *direct tests*.

#### **1-7-2-1 Indirect Tests**

They appeared earlier before the era of the communicative approach to testing. The test taker in this type of testing is not required to speak at all. O'loughlin (2001) mentions an example of this kind of tests, which was Lado's pronunciation tests which used to ask the examinee to point out the words that have different pronunciation from others.

#### **1-7-2-2 Direct Tests**

They are tests where the examinee participates in a face-to-face interaction with one or more interlocutors. The American Foreign Service Institute (FSI) introduced this type for the first time in the 1970s: the Oral Proficiency Interview (OPI). OPI is an oral unstructured interview between one examinee and an examiner who scores the performance. Direct tests were widely adopted since 1970s as the most relevant way of testing speaking skill proficiency (O'loughlin, 2001).

Direct tests or *live tests* are not as sufficient as to show the linguistic competence of the learner, nor are they practical in terms of time and they give an authority to the tester upon the

examinee. However, they have the advantage of giving more control to the tester over the learner's performance (Luoma, 2004).

Giving the drawbacks of interviewing only one examinee at a time, another way was introduced: interviewing pairs of learners without the interference of the examiner. This way, has three advantages. First, the interaction between two learners demonstrates their competence more than does an interaction between a teacher and a learner. Second, this way of testing simulates the way of teaching and thus the examinees feel as they are learning. Third, it is a practical way to gain time.

However, pair interview has its limitations. The influence of one partner on the other either by his or her personality or by his or her way and level of speaking may prevent the other from showing his or her real ability. The other point is that pair interview success is not always guaranteed; it may happen that one of them speaks a lot while the other speaks only a few words, or two of them speak little then stop. Hence, the examiner should be prepared to come out of this impasse (Luoma, 2004). "Direct tests are subjective and depend on whether the assessors are good at using grading rubrics/scales and test formats to minimize and reduce subjectivity and disagreement" (Weir, 1988, as cited in Al-Amri, 2010, p. 114).

Direct tests, also, were criticized by the lack of validity and reliability. As a response to this criticism, the University of Cambridge Local Examination Syndicate (UCLES) developed the IELTS exam (International English Language Testing System). It is a standardized test, which has the goal of testing the ability of learners to use English as a medium of instruction to learn. Testing speaking within IELTS is a structured interview consisting of five sections. First, the candidate and the interviewer introduce themselves. Next, the candidate is asked to talk extensively about a familiar topic. Then, he or she is required to solve a problem or to get an information. After that, a speculation task is performed, where the examiner speaks about his or her study or work plans. Finally, he or she concludes the conversation (UCLES, 1999, as cited in O'loughlin, 2001). Besides, UCLES developed other similar tests such as FCE (Cambridge First Certificate in English) and CPE (Certificate of Proficiency in English oral interviews). These task-based structured tests have gained a large popularity all over the world (O'loughlin, 2001).

### **1-7-2-3 Semi-Direct Tests**

They are performed without the presence of a human interlocutor. Clark explains that test takers in semi-direct tests complete the tasks using a tape-recorder or any non-human intermediary better than with a human interlocutor (1979, as cited in O'loughlin, 2001). A semi-

direct speaking performance testing is recorded then assessed by a trained assessor (O’loughlin, 2001). The main purpose of such tests is to maintain standardisation by getting the learners perform the same instructions. They have the advantage of being practical in terms of cost and time. However, they were criticised by being non-authentic because they do not simulate the real-life social interaction (Jeong et al, 2011). They were employed for the first time in the 1970s and became very popular in the last quarter of the twentieth century, mainly in the United States (O’loughlin, 2001).

Luoma (2004) refers to this type as a *Tape-Based Testing*. TSE (Test of Spoken English), ROPE (Recorded Oral Proficiency Examination), and SOPI (Simulated Oral Proficiency Interview) are examples of some American semi-direct speaking assessments. TEEP (Test in English for Educational Purposes), and AET (ARELS Examinations Trust of Oxford) are examples of British semi-direct speaking tests (O’loughlin, 2001).

From another perspective, Luoma (2004) distinguishes between two types of dichotomies “*Construct-based* versus *Task-based* assessment of speaking” and “*Stand-alone* versus *Integrated* assessment of speaking”.

#### **1-7-2-4 Construct-Based Versus Task-Based Assessment**

The emphasis of a construct-based assessment is primarily on the *construct* of ‘speaking ability’. In task-based assessment, the focus is on the ability of speaking in different social situations, especially in professional situations. Luoma mentions the difference between the two types

The difference between construct-based and task-based assessment is the position of tasks in designing assessments. Do they define what is tested, or are they the means of getting the examinees to talk so that their speaking skills can be assessed? The choice depends on the intended use of the scores (2004, p. 42)

The choice between construct and task based assessment relies on the purpose of the test designer. A construct-based assessment would be better when the purpose is to assess the general ability of speaking while a task-based choice is better when the goal of the tester is to show the ability of the learner to interact in different speaking contexts. It is worth noting that the two choices are not contradictory; rather they are complementary to each other (Luoma, 2004).

### 1-7-2-5 Stand-Alone Versus Integrated Assessment

Speaking assessments usually focus on the speaking skill and do not integrate the other skills, however other tests such as task-based tests do mix reading, listening and/or writing with the speaking skill. The purpose of mixing all skills in a speaking test is to create a real-life-like test. This type of testing is called *integrated assessments* while the former type is the *stand-alone* test (Luoma, 2004). The difference between the two perspectives occurs during the rating procedure. Raters, when rating the integrated tests focus on the “examinees’ comprehension of the input material and its effect on fluency and on quality and organisation of content” (Luoma, 2004, p. 44). Moreover, raters do not agree on the rating scale which generates a problem of reliability. This can be solved by agreeing on a detailed rating scale and engaging raters in training sessions (Brown et al, 2001, as cited in Luoma, 2004)

### 1-7-3 Speaking Assessment Tasks

Although it seems to be easy to arrange tasks for the assessment of speaking, the selection of tasks should be carefully studied because this is influenced by the purpose of the test and the ability to generalise the scores obtained to the use of the language outside the test. That is to say that tasks have to be authentic and relevant to the criteria being assessed (Fulcher, 2010). Valette states that:

Although a free expression test allows the students to demonstrate their linguistic creativity, one cannot simply put the student before a tape recorder or in front of the class and say ‘speak’. A testing framework that will facilitate student talk must be established (1977, as cited in Fulcher, 2010, p. 51)

Bachman and Palmer (1996) believe that in order to achieve the goals intended behind a test, the task characteristics included in this test have to be analogous to those of target language use (TLU) tasks.

Fulcher (2010, p. 57) introduces a framework that describes the characteristics of speaking test tasks. This framework includes six features: Task orientation, Interactional relationship, Goal orientation, Interlocutor status and familiarity, Topic, and Situation.

**Task orientation** can be open, guided, or closed. An open task means that the examinee is free to say what she or he wants. A guided task means that the examinee is guided by some

instructions but has a degree of flexibility, and a closed task means that the examinee is totally guided by rubrics. **Interactional relationship** refers to the discussion being occurred whether it is non-interactional, i.e. the examinee is the only speaker, or interactional, i.e. the existence of an interaction between the examiner and the examinee. The **goal orientation** may be divergent or convergent, or none of them. **Interlocutor status and familiarity** denotes the position of the interlocutor in the interaction. It can be higher such as a teacher, or lower such as a colleague, or none of them in case of a tape recording test for instance. **Topic** refers to the topic of the task if it is familiar or unfamiliar (Fulcher, 2010).

Brown and Yule (1983, as cited in Luoma, 2004) sort four types of talk according to their degree of difficulty: description, instruction, storytelling and opinion expressing and justification. Another viewpoint organise speaking tasks to *functions* and *notions*. The former refers to what learners need to be able to do, and the latter refers to what they need to be able to talk about (van Ek, 1975; and Wilkins, 1976, as cited in Luoma, 2004).

Luoma (2004) classifies speaking tasks into three types: *open-ended*, *semi-structured*, and *structured* tasks. *Open-ended tasks* direct the talk, but give opportunity to demonstrate the language competence through performing long speaking. Tasks included in this type are description, narrative, instruction, comparison, explanation, justification, prediction, and decision tasks. The materials used can be written scripts, pictures, or an instruction from the examiner. Tasks of this type can be interactive or monologue. Interactive means that there is an interaction between two or more people such as role-play and discussion. The monologue is a one-direction speaking task, such as presentation, lecture, and speech.

*Structured tasks* limit the choice of speakers on what to say or even provide them with a list to choose what to say. Even though these tasks are not useful for assessing the speaking creativity of learners, teachers are objective in scoring learners' performances and thus are able to make a fair comparison between them. Structured tasks such as reading aloud, sentence repetition puts emphasis on pronunciation and grammar knowledge while short-answer questions and reactions to phrases (where the examinee is given the first turn of speaking and is required to give the second one) tend to elicit the comprehension of context. *Semi-structured tasks* focus on social and functional use of the language. The learners are required to imagine themselves in a given situation and asked to express what would they say or how would they react. This type of tasks fits more tape-based tests because it is difficult for a tester to change roles many times.

Brown (2004, p. 141) introduces five types of assessment tasks: imitative, intensive, responsive, interactive, and extensive tasks.

**Imitative speaking Tasks** involve imitating tester's set of words or sentence. Although imitation is a phonetic task, its purpose in a speaking test is purely to show the test-taker's pronunciation.

**Intensive speaking Tasks** contain producing short utterances in order to show the test taker's competence in a precise area of language such as grammar, phonology, or vocabulary. In these tasks, it is necessary for the test taker to have a sufficient knowledge of meanings in order to generate the correct response. However, interaction is little between the speaker and the interlocutor because the response is controlled and likely to be mechanical.

Tasks such as direct response, reading aloud, sentence and dialogue completion, and limited picture-cued tasks are examples of intensive speaking tasks. *Directed Response Task (DRT)* is used to bring out precise forms of grammar, or to transform sentences. It is not a communicative task for it requires a limited mechanical response. Brown provides an example of this task:

*Test takers hear:*

*Tell me he went home.*  
*Tell me that you like rock music.*  
*Tell me that you aren't interested in tennis.*  
*Tell him to come to my office at noon.*  
*Remind him what time it is.*

(2004, p. 147)

*Reading Aloud Task* requires the test taker to read aloud a selected passage while the tester is recording. Testers after listening to the recorded reading give a score relying on a number of phonological criteria such as stress and intonation. This technique is used in order to get an overview of the examinee's general oral production skills. Some past speaking tests used this technique using a scale for pronunciation and fluency. However, it is criticized by being inauthentic and far from showing the real communicative ability.

In a *Sentence and Dialogue Completion Task*, the test taker is given a written dialogue between two persons in which the response of one of them is omitted. The examinee has some time to read the dialogue and expect what would be the omitted response, and then play the role with the tester. *As for an Oral Questionnaire*, the examinee is given a written questionnaire and is required to answer the questions orally. The purpose of such tasks is to shed the light on the test taker's sociolinguistic oral ability.



A *Picture-Cued Task* is a task that demands the test taker to look at some pictures and describe what he or she is looking at. The description varies from close-ended responses such as providing some vocabularies (names, numbers, colours, describing people) to open-ended descriptions such as storytelling or describing directions and locations.

**Responsive Speaking Tasks** have the goal of demonstrating the interactional ability of the examinee. They comprise interactions that are not as short as the intensive tasks, nor are they long as interactive tasks. Examples of this type of tasks encompass *question and answer*, *giving instructions and directions*, and *paraphrasing*. *Question and Answer Tasks* are administered in a form of interview, which includes few oral questions addressed to the test taker who has to answer them.

Two kinds of questions are used: *display* questions, and *referential* questions. The former type requires the test taker to provide predetermined right responses, while the latter one offers him or her the opportunity of producing expressive utterances. *Giving instructions and directions* is a useful task that encourages the learner engaging in a somewhat long speech. The assessor has to prepare in advance the type of questions whose responses include instructions such as ‘how to prepare a traditional dish’, ‘how to drive a car’, etc. The purpose of such tasks is to get the speaker’s instructions.

It is worth noting that test takers must not know the questions before the day of the assessment lest they prepare the answers and just parrot them back to the examiner. Another important point is that the topic should be simple and reachable by anyone because the purpose is not testing the general knowledge about that topic, but testing the linguistic competence of the learners. *Paraphrasing* is to listen to and understand some sentences and to produce other sentences that are different from the heard ones but have the same meaning as them. An authentic example of this is to listen to a call message, then transfer this message to another person.

**Interactive Speaking Tasks** include interaction that is longer and more complex than the above-cited tasks. Interaction can be *interpersonal* or *transactional*. Interpersonal interaction involves different sociolinguistic conventions such as humour, slang, and ellipsis. It aims at keeping social relationships. Transactional interaction aims at delivering special information. Interviews and role-plays are examples of interactive tasks.

**Extensive Speaking Tasks (Monologue)** contain less or no interaction. Speech, storytelling, oral presentation are examples of this type of tasks.

### 1-7-4 Scoring Speaking Tests

One of the most important decisions in speaking tests is to transform the examinee's performance into scores (Brown, 2004). Speaking scores are used to reflect the test taker's ability to do a given test task. They can be numbers or letters usually accompanied with short expressions explaining the meaning of each one (Luoma, 2004). To do this, testers, usually, use a rating scale "sometimes referred to as a 'scoring rubric' or 'proficiency scale'" (Fulcher, 2003, p. 88). A rating scale is defined as:

.....Consisting of a series of constructed levels against which a language learner's performance is judged... Such scales range from zero mastery through to an end-point representing the well-educated native speaker. The levels or bands are commonly characterised in terms of what subjects can do with the language (tasks and functions which can be performed) and their mastery of linguistic features (such as vocabulary, syntax, fluency and cohesion)

(Davies et al., 1999, as cited in Fulcher, 2003, p. 89)

However, when developing a rating scale, it is a challenge to reflect a so complicated phenomenon such as speaking performance in just a number or a word relying on an imperfect theory (North, 1996, as cited in Luoma, 2004). Moreover, the description of each scale is purely human. This difference reflects different assumptions and thoughts about learners' achievements; consequently, it reflects different definitions of the constructs to be assessed (Brindley, 1998, as cited in Luoma, 2004). Gronlund (1998) suggests that grading has to be based on,

The student achievement, and achievement only. Grades should represent the extent to which the intended learning outcomes were achieved by students. They should not be contaminated by student effort, tardiness, misbehaviour, and other extraneous factors....If they are permitted to become part of the grade, the meaning of the grade as an indicator of achievement is lost

(As cited in Brown, 2004, p. 284)

Nevertheless, other viewpoints do not agree with this idea. They advocate the idea of triangulation in order to know the learners' abilities, because it is difficult to demonstrate these

abilities using only the achievement tests and measured performances. They suggest that examinees' struggles, enthusiasm, behaviour, attendance and progress can be criteria that are added to the final grade (Brown, 2004).

It is worth noting that rating scales differ regarding the purpose they are set for, and the targeted people (Alderson, 1991; North, 1996; as cited in Luoma, 2004). In this vein, Alderson mentions three types of scales: *user-oriented*, *assessor-oriented* and *constructor-oriented* scales. *User-oriented* are related to what the examinee is able to do in each level. *Assessor-oriented* scales focus on the definition of the assessed construct. *Constructor-oriented* scales help the tester to select the relevant tasks for the test that are meant to elicit the required speaking sample (Alderson, 1991b, as cited in Fulcher, 2003). Yet, the Council of Europe refers to these scales as *examinee-oriented*, *rater-oriented*, and *administrator-oriented*:

Rater-oriented scales must help raters make consistent decisions; examinee-oriented scales give information about overall level and, possibly, particular strengths and weaknesses, while administrator-oriented scales give overall information in a concise form (2001, as cited in Luoma, 2004, p. 60)

Scoring speaking performance might be holistic or analytic. *Holistic scoring* refers to evaluating the overall speaking ability of the examinee. It is a quick way of evaluation, and is useful in case of informal appraisals (Thornbury, 2005). It is considered as practical in terms of decision-making because the rater has to afford only one mark, and flexible in that they "allow many different combinations of strengths and weaknesses within a level" (Luoma, 2004, p. 62). However, it is not practical in determining what area needs improvement and which area does not, and in making a clear distinction between a level and another especially when using quantifiers such as *well*, *enough*, *good*. (Luoma, 2004). *Analytic scoring* however relies on many selected criteria each of which is scored independently and the final mark is the sum of these scores. It is more reliable than the holistic scoring (Thornbury, 2005), and it provides a clear insight over the strengths and weaknesses of the learners' performances (Luoma, 2004).

Hamp-Lyons (1991, as cited in Fulcher, 2003) divided the holistic assessment into three types of scoring: *holistic scoring*, *primary-trait scoring*, and *multiple-trait scoring*. *Holistic scoring* is to offer a single grade to the learner's overall speaking performance. The limitation of this kind of scoring is in that it deals with 'speaking' as a whole phenomenon and not with its sub-skills as constructs, thus it lacks fairness in some cases. *Primary-trait scoring* relies on

the assumption that each speaking performance has to be assessed in its context and the rating criteria used in this assessment is relevant only for this context. **Multiple-trait scoring** assigns a score for each trait or each construct of the performance. This scoring type is useful to determine the weaknesses and the strengths of a performance. Below are some examples of rating scales:

- **The Interagency Language Roundtable (ILR) rating scale** appeared in the USA in 1991 after the FSI rating scale. The ILR is a holistic scale that assesses the examinees' performances against a range of constructs definitions with reference to the tasks that the examinee is able to do in natural situations. Bachman refers to this scale as a '*real-world*' approach, while he refers to his own scale as an '*ability/interactional*' approach (Fulcher, 2003).
- **The Cambridge Certificate in English Language Speaking 'CELS'** test relies on an analytic rating scale that includes grammar, vocabulary, discourse management, pronunciation, and interactive communication (Thornbury, 2005).
- **The National Certificate scale** is a Finnish National Certificate holistic scale that contains six levels varying from the most advanced speaker to the early beginner one. These levels are checked through three tests, each one includes two levels (Luoma, 2004).
- **The ACTFL** (American Council for the Teaching of Foreign Languages) speaking scale is similarly a holistic scale that encompasses four levels: *superior*, *advanced*, *intermediate*, and *novice* levels. Each level describes what can the learner do using the language in different contexts (Tuttle & Tuttle, 2013). The ACTFL scale is used to show the progress of learners specifically in US educational contexts, that is why it is criticized as being a behavioural approach rather than a theory based approach. That is to say, that it describes the learners' language use in a particular situation instead of their theoretical knowledge according to a model of communicative competence. The advantage of this scale is that it attracts the American teachers' interest in language use rather than language knowledge (Luoma, 2004).
- **The IELTS speaking** rating scale involves "fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation" (Hughes 2011, as cited in Pawlak & Waniek-Klimczak, 2015, p. 253).
- **The Test of Spoken English (TSE) scale** is a mixture of analytic and holistic scoring, and is a theory-based approach to scoring consisting of five levels. The version of TSE

addressed to the examinee encompasses functional competence, sociolinguistic competence, discourse competence and linguistic competence, while the version addressed to the rater is more detailed and contains descriptions of what the examinee is able to do in each level (Luoma, 2004).

- **The Common European Framework** issued by the council of Europe in 2001 is a scale that is not intended for a particular test. It describes the learners' abilities to do different skills at different levels. The CEFR is a measurement approach that helps learners to check their learning levels (see appendix 3) (Fulcher, 2003).

### 1-7-5 Applying Principles to Speaking Tests

Regarding the complexity of the speaking skill, test designers have to follow specific principles to guarantee effective speaking tests. These principles are reliability, validity, practicality, authenticity, and washback.

Reliability of speaking tests is achieved when scores are similar on different occasions. This can be achieved by using the same rating scales. Rater training is the means that helps raters to use similar rating scales (Al-Amri, 2010). Reliability of scores “*means that the scores are dependable, so that we can rely on them in decision-making*” (Luoma, 2004, p. 176). Reliability of scores can be achieved whenever a consensus upon a consistent rating process is made between raters; however, this is difficult giving the subjectivity of human beings and the different thoughts and techniques they follow to score test performances. One useful way to alleviate this problem and guarantee reliable scores is ‘*rater training*’ during which raters learn to distinguish between different levels of abilities using concrete terms and analytic scoring. Another way to improve score reliability is to set ‘cut scores’ (Luoma, 2004). Cut scores are

..... Selected points on the score scale of a test. The points are used to determine whether a particular test score is sufficient for some purpose. For example, student performance on a test may be classified into one of several categories such as basic, proficient, or advanced on the basis of cut scores.... Cut scores should be based on a generally accepted methodology and reflect the judgments of qualified people (Zieky and Perie, 2006, p. 2)

The implementation of cut scores is not an easy task that any examiner can do it rather it requires the interference of experienced test designers, educators, policymakers and other experts (Zieky and Perie, 2006).

A third way that is likely to ensure the reliability of test scores is ensuring the stability of rating procedures by using the '*rating forms*'. Rating forms are scripts where raters record their rating process. Raters, often, allot to each examinee a rating form that mentions every single information concerning the rating process such as the examiner and the examinee's names, the tasks performed, the rating criteria, and the remarks of the rater if existing. Raters should agree upon a shared rating process. The more the criteria are well defined the more the subjectivity is avoided and the scores are reliable (Luoma, 2004).

Concerning validity, few studies have discussed the validity of speaking tests (Sak, 2008). Luoma (2004) states that for a speaking test validity, the meaningfulness of scores is the corner stone. Scores validity reflects many interesting matters in speaking tests. The purpose of the test is the first matter that scores should serve. To do this, test developers have to provide a clear and precise definition for the constructs they intend to assess. There are three approaches to defining the test constructs the *linguistic approach*, the *communicative approach*, and the *task-based approach*. **The linguistic approach** concentrates only on the appearance of the language such as grammar, vocabulary, pronunciation, and fluency. **The communicative approach** sheds the light on the ability of the examinee to use the skills that the task demands. **The task-based approach** attempts to simulate real-situations and asks the examinee to take a given role, such as the role of a patient at a clinic.

Given that the purpose of the test has been determined and the constructs have been defined, test developers have to demonstrate the correlation between the test purpose, the construct definition, and the designed tasks. Likely, rating criteria should be relevant to the test purpose, the definition of the construct being assessed, and the test tasks. The administration and the rating procedures involve reports coming from administrators and raters, rater training materials, interlocutor and rater qualification (Luoma, 2004). These matters "can be evaluated in terms of their consistency and their coherence with the construct definition" (Luoma, 2004, p. 186)

Al-Amri considers that spoken English tests "are valid when they are able to give a clear idea about the ability of the candidates to communicate, for example, in real life situations" (Al-

Amri, 2010, p. 113). Al-Amri believes that the face-to-face interview is an instance where the examiner can know about the ability of the examinee of speaking in authentic situations.

The issue of reliability and validity within speaking tests is not as simple as within other tests regarding the complexity of measuring the speaking skill. This did not prevent some researchers to embark studies highlighting this issue. Nakamura (1997, as cited in Sak, 2008) studied the construct validity of a speaking test of English. The purpose of his research is to know the relevance and separateness of nine sub-constructs (*pronunciation, grammar, discourse, fluency, content, vocabulary, comprehensibility, interactional competence and sociolinguistic competence*) of speaking to speaking ability. The result from this study shows that,

The nine traits were functioning as factor construct elements in both direct oral communication ability (Interview test) and semi direct oral communication ability (tape mediated test). However, they all maintain their own characteristics, which cannot be measured by others (Sak, 2008, p. 28)

Iwashita et al (2007, as cited in Sak, 2008) studied the validation of test scores through using rating scales for a new international test of English for academic purposes. The researchers used five features in the rating scale; grammatical accuracy, complexity, vocabulary, pronunciation and fluency. The study showed that vocabulary and fluency had the strongest influence on the final grade. It showed also that the weakness of the examinee in one feature does not influence his or her overall proficiency. These results are likely to play an important role in the scale development and thus in validating test scores.

The study of Salabary (2000, as cited in Sak, 2008) also investigated the validity and reliability of the OPI (Oral Proficiency interview) test in terms of tasks and rating criteria. He advocated the choice of conversational interaction tasks such as role-plays as they are more authentic.

As regards practicality, Luoma (2004) considers it as a decisive factor for speaking tests. If the test requirements are highly costing in terms of materials, money, and administrators, it may be rejected (Luoma, 2004). Face-to-face speaking tests are not practical because they consume much time to gather data about the examinees' speaking ability and if testers try to

perform them in a short time, the validity of the test will be influenced. To reduce this problem, a training on the use of criteria and the rating scales remains very important (Al-Amri, 2010).

In addition to what has been said, test developers have to pay attention to the examinees' attitudes to the tests and the impact of the test on their learnability and on the teaching process (Luoma, 2004). Ur (1996) stresses that the speaking skill is important and should take the priority in language exams because,

If you have an oral proficiency test at the end of a course, then this will have a “backwash effect”: teachers and students will spend more time on developing skills during the course itself. Conversely, if you do not have such a test they will tend to neglect them. Students who speak well but write badly will be discriminated against if all or most of the test is based on writing (As cited in Sak, 2008, p. 14)

Likewise, Kitao and Kitao (1996, as cited in Trejos et al, 2013) ascertain that speaking tests have a beneficial washback in that they enhance the teaching of this skill in the classroom. Hughes (1989) states that “if you want to encourage oral ability, then test oral ability” (as cited in Knight, 1992, p. 294). Hughes means that a test of oral ability should have a positive impact in promoting the oral ability of students.

## **1-8 Conclusion**

This chapter has dealt with assessment as a general concept and with the assessment of the speaking skill specifically. It has shed the light on the main concepts included in assessment such as testing, evaluation, formal, informal, achievement, proficiency, performance, summative, formative, and so on. Then, it has given a summarized history of tests since its beginning until the recent years where the communicative approach is dominating the testing field. After that, the main concept of this research work ‘speaking’ has been discussed in a some detailed way. It has first dealt with the nature of this skill and what has been said about it. Then, it has moved to its assessment taking into consideration the most issues related to it such as the types of speaking tests, the challenges faced when designing and administering speaking tests, the tasks often introduced to meet the purpose of such tests, the scoring process, and the principles that should characterize these tests including reliability, validity, practicality, and washback.



The next chapter will get a look over the practical side of this research work. It will introduce the educational situation in the Algerian Universities with close reference to the University of Tlemcen and the section of English within it. It will also get the reader know the instruments used to collect data about the assessment of the speaking skill and the way these instruments were exploited.

# CHAPTER TWO

## DATA COLLECTION

# CHAPTER TWO

## DATA COLLECTION

<b>2-1 INTRODUCTION.....</b>	<b>36</b>
<b>2-2 THE ALGERIAN EDUCATIONAL SITUATION.....</b>	<b>36</b>
2-2-1 ENGLISH LANGUAGE TEACHING AT THE MIDDLE LEVEL.....	37
2-2-2 ENGLISH LANGUAGE TEACHING AT THE SECONDARY LEVEL.....	37
2-2-3 ENGLISH LANGUAGE TEACHING AT THE UNIVERSITY LEVEL.....	38
<b>2-3 RESEARCH METHODOLOGY.....</b>	<b>42</b>
2-3-1 QUANTITATIVE RESEARCH APPROACHES.....	42
2-3-1-1 DESCRIPTIVE SURVEY RESEARCH.....	43
2-3-1-2 EXPERIMENTAL RESEARCH.....	43
2-3-1-3 CAUSAL-COMPARATIVE RESEARCH.....	43
2-3-1-4 CORRELATIONAL RESEARCH.....	43
2-3-1-5 META-ANALYSIS RESEARCH.....	43
2-3-2 QUALITATIVE RESEARCH APPROACHES.....	43
2-3-2-1 ETHNOGRAPHIC STUDY.....	44
2-3-2-2 GROUNDED THEORY RESEARCH.....	44
2-3-2-3 PHENOMENOLOGY RESEARCH.....	44
2-3-2-4 CASE STUDY.....	44
<b>2-4 RESEARCH DESIGN.....</b>	<b>45</b>
2-4-1 RESEARCH OBJECTIVES.....	45
2-4-2 SAMPLING.....	46
2-4-3 DATA COLLECTION.....	48
2-4-3-1 THE OBSERVATION.....	49
2-4-3-2 THE STUDENTS' QUESTIONNAIRE.....	50
2-4-3-3 THE TEACHERS' INTERVIEW.....	51
<b>2-5 CONCLUSION.....</b>	<b>53</b>

## **Chapter Two: Data Collection**

### **2-1 Introduction**

This chapter is offered to describe the educational situation at the University of Tlemcen and to shed the light on the practical side of this research work.

The researcher elucidates the different steps that are followed to gather data about the assessment of the speaking skill within the English department at the University of Tlemcen. The researcher begins with the methodology of the research, then shifts to the research design that includes the research objectives, the tools used in collecting data, the targeted population and the sample chosen from this population, and finally describes in detail the procedures of each instrument.

### **2-2 The Algerian Educational Situation**

With the change occurring in almost all the domains, and the continuing process of globalisation, the English language shifted from an international to a necessary global language of interaction between countries. Algeria was not an exception. Reforms became inevitable, namely the educational reforms. The policy makers accredited the CBA approach to teaching at the middle and secondary levels (Benmostefa, 2014).

The CBA or Competency Based Approach was introduced in Algeria in July 2002. It is based on the premise that the teacher is no longer the centre of the classroom. Learners are competent enough to be responsible of their learning process and capable of understanding the world that surrounds them by using the skills they acquired to solve real-life problems (Aimeur, 2011). The competency-based approach enables the learner to build academic and life skills such as how to communicate in different situations, and how to interact socially and emotionally (Chelli, 2010)

Talking about ELT, the Algerian Ministry of Education distinguishes two main aims of English teaching in Algeria:

To increase learners' cross communicative skills and develop the exchange of ideas across nations and to develop their ability to have access to materials

written in English as far as graduate or/ and post-graduate students are concerned (Pedagogical Instruction, 2004, as cited in Berrabah, 2006, p. 23)

English is planned to be a bridge of communication between the Algerian and the international learners. Accordingly, the teaching of English language adopted the CBA approach at the middle and the secondary level.

### **2-2-1 English Language Teaching at the Middle Level**

The introduction of the CBA required the implementation of a new programme and new textbooks that stand for it. In the academic year 2002/2003, the Algerian Ministry of Education supplied the four levels of the middle school with new textbooks. “*Spotlight on English*” for the first year, “*The Second English Course book*” for the second year, “*Spotlight on English*” for the third year, and “*On the Move*” for the fourth year. Through the new syllabus, the Ministry aimed at having learners grasp the basic knowledge of the English language and adapting them to the teamwork and group discussion through projects. This would foster their communicative competency (Djebbari, 2014).

Four objectives were namely considered at this level: linguistic, methodological, cultural, and socio-professional objectives. From a Linguistic angle, the new programme aims at developing the learners’ grammar knowledge. It aims also at broadening their range of vocabulary and providing more training in the four skills to enhance their communicative and interaction competencies. From a methodological standpoint, the syllabus has an objective of boosting the learner’s autonomy by teaching them how to work and think critically. Similarly, the programme did not neglect the cultural side of language learning. It designs several tasks that get the students discover the culture and the civilization of the English people (Khelifi, 2014).

### **2-2-2 English Language Teaching at the secondary Level**

The secondary level, which consists of three years and ends up with the Baccalaureate exam, had been also a subject for the educational reforms in 2006. Under this reform, all the students in the first year are enrolled in either a scientific or a literary stream. English is included in the two streams. The programme of English in this year is similar to that of the middle school. It involves grammar, vocabulary, and study skills. Learners are supposed to have an efficient communication, an adequate grammar, and learning autonomy. These objectives seem to be

important and ambitious, but unreasonable and unachievable, because there are some factors obstructing the realization of them such as time constraint, the overloaded curriculum, and the low motivation of the pupils. Concerning the second year, the main objective is to develop communication by having the learners to get used to the teamwork, to know the strategies of communication, and the language structures. As for the third secondary year, the learners are supposed to be familiar with the necessary communication skills, and thus the focus shifts to the study of discourse patterns and language functions. However, the overloaded curriculum planned for this objective remains beyond the capacity of learners and teachers alike (Bouyakoub, 2012).

The implementation of CBA faced many barriers when it comes to practice because either the teachers were not trained on how to put such an approach into practice, or the approach was not relevant to the Algerian context. Consequently, the majority of them ended up following their old teaching methods (Djabbari, 2014). Along with the reform movements, the Algerian Ministry of Higher Education and Scientific Research introduced the LMD system to the university level.

### **2-2-3 English Language Teaching at the University Level**

English is taught almost in all the Algerian universities either as a major at the language departments, or as a module (ESP) at the other departments (such as the scientific and technologic departments). Prior the 1980s, learners had to spend three years to get a degree (License) in the “*English language and its literatures*”. By the mid-eighties, and with the advent of the communicative approach, the duration required to get this degree had been changed to four years. The first year aimed at strengthening the basics of the language previously acquired. The modules taught at this level were grammar, written expression, reading comprehension, oral expression, phonetics, linguistics, and Arabic. These modules were taught until the second year excluding the oral expression, which was retained to the third year, however modules such as literature and civilisation were taught in the second, third, and the fourth year. This learning system stayed until 2000s, when new reforms were introduced to the higher education (Bouyakoub, 2012).

The LMD system includes three cycles: the first cycle is an undergraduate study, which lasts six semesters (three study years). The students are awarded the *License* degree at the end of this cycle. The second and the third cycles are postgraduate studies. The second cycle goes

on four semesters, the equivalence of two study years, and is crowned by the Master degree, while the third cycle, sustaining six semesters, (which corresponds to three years) is finished off by getting the *Doctorate* degree (Djebbari, 2014).

The implementation of the LMD system in Algeria aims at keeping abreast with the ongoing globalisation process (Djebbari, 2014). The LMD system's main objective is

Bridging the gap between university formations, training and theoretical qualifications and the outer, practical and professional world of work and labour is the rationale behind such an endeavour of engaging a whole system of education through the LMD system (Berrabah, 2014, p. 162)

The LMD system is meant to modernize the Algerian universities, and to provide it with the latest development in the education field. Regarding ELT field, the adoption of the LMD system is intended to improve the learner's capacities in using English to express themselves either orally or by writing and to enable them to acquire a knowledge about the literature and the civilisation of the English people in order to build a full image of the English Language. Each semester includes a number of teaching units combining different subjects depending on the level of study and the specialty followed. During all the six semesters of the first cycle four types of units are involved fundamental, methodology, cross-sectional and discovery unit (Khaldi, 2014). The two tables below clarify the programme of the six semesters:

		Modules	1 <sup>st</sup> & 2 <sup>nd</sup> Semester (hours/week)	3 <sup>rd</sup> & 4 <sup>th</sup> Semester (hours/week)
<i>Fundamental Teaching Units</i>	Competence in English Language	Techniques of Oral Production	3	3
		Techniques of Written Production	3	3
		Discourse comprehension	3	3
	Description of the Language	Grammar	6	6
		Phonetics	3	3
		Introduction to linguistics	1.5	1.5
	English Culture	Introduction to literary studies	1.5	1.5
		Anglo- Saxon Culture and civilization	1.5	1.5
		African civilization	/	1.5
			/	
<i>Methodology Teaching Unit</i>	Research Methodology	1.5	1.5	
<i>Cross-sectional Teaching Unit</i>	Information and Communication Technologies (ICTs)	1.5	1.5	

*Table 2.1: The programme of the first four semesters*

What is clearly observable in this table is the absence of the listening module in either the first or the second year. In fact, it is incorporated with the oral production module



in order to provide a model of speaking to the students to imitate, and thus they can be able to grasp the speech pace, intonation, and stress. (Khaldi, 2014).

5 <sup>th</sup> & 6 <sup>th</sup> Semester (hours/week)					
Language Studies			Literature and Civilisation		
<i>Fundamental Teaching Units</i>					
<b>Description of the Language</b>	Theories of linguistics	3	<b>Literary Culture</b>	Theories of Literature	3
	Phonology	3		Comparative Literature	3
	Morphology	3		Literature Culture	3
<b>Application of the Language</b>	Psycholinguistics	3	<b>English Cultures</b>	English Cultures	3
	Sociolinguistics	3		Anglo-Saxon Civilisation	3
	TEFL	3		African Civilisation	3
<i>Methodology Teaching Unit</i>					
	Research Methodology	3		Research Methodology	3
<i>Discovery Teaching Unit</i>					
	Project / Training	3		Project / Training	3
<i>Cross-sectional Teaching Unit</i>					
	Educational psychology	3		Educational Psychology	3

**Table 2.2: The programme of the fifth and the sixth semesters**

Since the focus of this research paper is on the speaking skill, one can remark the low time allotted to this skill; three hours per week in each semester. However, it is not figured at all in the programme neither of the fifth nor of the sixth semester. On the other hand, modules such as phonetics and phonology standing up to reinforce the speaking skill in an indirect way are both assigned three hours per week.

## 2-3 Research Methodology

A research is a way to uncover the truth (Cohen et al, 2007); an organised enquiry process that includes forming questions and hypotheses, collecting data, analysing and interpreting these data (Nunan, 1992). Kothari defines research as a “search for knowledge..., a scientific and systematic search for pertinent information on a specific topic. In fact, research is an art of scientific investigation” (2004, p. 1). In other words, a research is a controlled and logical way to look for reality.

Research, within the field of education is “a disciplined attempt to address questions or solve problems through the collection and analysis of primary data for the purpose of description, explanation, generalization and prediction” (Anderson, 1998, p. 6). Educational researches are classified into two categories: *quantitative* and *qualitative* research. Quantitative research relies on statistics and quantification whereas Qualitative research does not rely on numbers; rather, it investigates phenomena in terms of quality (Kothari, 2004). Quantitative research, also called *the scientific approach*, assumes the social reality as being objective. However, the qualitative, or *the naturalistic approach* to research, does not neglect the subjectivity of individuals and considers the quality of subject matters (Blaxter, 2006).

In recent times, prevalent discussions have taken place concerning the adoption of either the quantitative or the qualitative approach in research. Some voices see that research should be either qualitative or quantitative, while others advocate the idea of mixing both of (Blaxter, 2006). Lodico & al (2006) note that some researchers consider that the Mixed-method research is likely to help comprehending more the problems of education.

### 2-3-1 Quantitative research approaches

Lodico & al (2006) illustrate that the feature distinguishing all quantitative approaches to research is the generation of statistical analyses. Bell claims that “Quantitative researchers collect facts and study the relationship of one set of facts to another. They use techniques that are likely to produce quantified and, if possible, generalizable conclusions” (2005, p. 8)

There exists many types of quantitative approaches: Descriptive Survey Research, Experimental Research, Causal-Comparative Research, Correlational Research, and Meta-Analysis.

**2-3-1-1 Descriptive survey research**

This approach has the aim of describing people's comportments, insights, thoughts, and principles about an existing subject in education. Data are collected through structured interviews, questionnaires, and tests (Scott & Robin 2011). Respondents are chosen randomly and results are categorised and generalised in a percentage form (Lodico et al, 2006).

**2-3-1-2 Experimental research**

The experimental research, usually, is thought of as related to the physical sciences. Nevertheless, experiments are relevant to the social science researches that have the purpose of identifying causal relationships such as psychology, and education, (Blaxter, 2006). Experiments are conducted to experience the degree to which different variables are related to each other (Nunan, 1992).

**2-3-1-3 Causal-comparative research**

Is also called **ex post facto**, and is the research that seeks what causes the differences between two groups, which have similar conditions but different experiences (Lodico et al, 2006).

**2-3-1-4 Correlational research**

This approach has the purpose of knowing the relationship between two measurable variables (Jackson, 2009).

**2-3-1-5 Meta-analysis Research**

This kind of research works as a method of study research by which researchers collect data through summarizing the findings of a number of chosen empirical studies (Lipsey and Wilson, 2001).

**2-3-2 Qualitative Research Approaches**

Qualitative approaches are categorised by their narrative findings. Bell states that:

Researchers adopting a qualitative perspective are more concerned to understand individuals' perceptions of the world. They seek insights rather than statistical perceptions of the world. They doubt whether social 'facts' exist and question whether a 'scientific' approach can be used when dealing with human beings. Yet there are occasions when qualitative researchers draw on quantitative techniques, and vice versa (2005, p. 8)

Case Study, Ethnographic Study, Grounded Theory, Phenomenological Study are types belonging to the qualitative approach.

### **2-3-2-1 Ethnographic study**

A study that looks into people in their natural location. Data are concerned with their social order and usual tasks (Brewer, 2000, as cited in Bell, 2005). Ethnography is the science that describes the culture of communities (Lodico & al, 2006).

### **2-3-2-2 Grounded Theory Research**

Is the approach where theory is generated when the data are being analysed; that is to say, that data are the ground for a new theory (Bell, 2005).

### **2-3-2-3 Phenomenology Research**

Phenomenology looks at the human phenomenon and its nature to describe it as free from any external influence (Cohen, 2007).

### **2-3-2-4 Case Study**

A case study is an enquiry that seeks to answer particular questions about a subject relying on various types of evidence (Gillham, 2000). Yin (2003) defines the case study as being the experimental investigation of a current phenomenon that relies on multiple sources of information brought from primary and secondary sources. Anderson (1998) states that the most important concern of a case study is how things occur and why.

The case study approach enables the researcher to undertake an investigation that cannot be undertaken by other research approaches, discover the complexity of social phenomena, explore the hidden reality, and see the case from the bottom-up, that is to know the viewpoints of the participants (Gillham, 2000).

Case studies are classified into three types according to their main purposes: exploratory, descriptive, or explanatory (Yin, 2003). The exploratory case study generates questions or hypotheses for further studies. The descriptive case study is concerned with describing the phenomenon within its real life context; the explanatory case study explains how events occurred (Yin, 2003). However, Merriam (1988, as cited in Cohen et al, 2007) provides another classification for case studies: descriptive, interpretive, and evaluative. Descriptive case studies provide narrative interpretations. Interpretive ones rely on the intuition of the researcher. Evaluative case studies focus on the judgement of the issue under investigation.

The present research work is an explanatory case study which enables the researcher to get a deep knowledge about the subject matter.

## 2-4 Research Design

As it was mentioned before, this research work is a case study. According to Yin (2003, p. 21), a case study design consists of five components: a study's questions, its propositions, its unit(s) of analysis, linking the data to the propositions, and the criteria for interpreting the findings.

**The study questions** are to be clarified precisely as the most important step.

**Study propositions** are the thoughts put by the researcher as answers to the questions. The researcher has to read the convenient literature and form the appropriate propositions before collecting any data (Yazan, 2015).

**The Unit of analysis** is related to the case to be studied. It can be individual such as a student, or multiple such as a group of people.

**Linking data to propositions** means matching the evidence found through analysing data to the suggested hypotheses (propositions) to see the degree of fitness between them (Heck, 200).

**Criteria for interpreting the findings** are the criteria that show the degree to which a research is successful (Griffie, 2012). Lodico et al (2006) suggest a design that includes determining the principal research topic, then embarking on a literature review; this helps forming or reforming the research questions. Next, the researcher has to consider his or her role in the study, the more the researcher is closer to the participants the more he or she gets in-depth data. After that, he or she has to identify where the study will take place and choosing the relevant participants. These steps are important to generate the study questions and start gathering data using several tools such as observations, interviews, and document analyses. The gathered data have to be analysed and interpreted to pick out the results that represent the findings of the research.

### 2-4-1 Research Objectives

The major aim of this research work is to investigate the assessment of the second year students' speaking performance at the University of Tlemcen. To do this, many sub-aims are necessary to be looked into. The first one is to know the way the speaking skill is taught and learnt since this may influence the students' performance during a test. The difficulties encountered either in teaching or in testing students is another sub-aim in this study. A third objective is the assessment types used by teachers to reveal the real students' abilities, and to probe whether these types are beneficial to motivate students. A fourth one is the way the

teachers rate their students' performances. Lastly, this study aims at giving some suggestions posed by the students and teachers to improving the students' speaking skill.

### 2-4-2 Sampling

The sample is "the group of elements, or a single element, from which Sample: Group of subjects from data are obtained" (McMillan, 1996, p. 85). According to Cohen et al (2007), the sampling strategy plays a crucial role in determining the quality of a research work. Sampling means focusing on a small group that represents a part of a bigger group. This small group is termed 'sample'.

Blaxter (2006) points out that the sampling strategies fall under two main categories, *Probability* and *Non-Probability* sampling. Probability sampling approach encompasses five types: *simple random*, *systematic*, *stratified*, *cluster* and *stage* sampling. Non-probability sampling approach involves six types: *convenience*, *voluntary*, *quota*, *purposive*, *dimensional*, and *snowball* sampling.

Probability sampling "is the gold standard of sampling.....a method used in research whereby any person in the population has a specified probability of being included in the sample" (Beins & McCarthy, 2012, p. 97). Among probability sampling strategies, *simple random sampling* is, perhaps, the most widely used type where everyone in the whole population has the same chance of being selected. *Systematic sampling* refers to the selection of one case every *n*th (a determined number) cases (Blaxter, 2006). *Stratified sampling* is the process by which the researcher divides a population under question into groups and selects a member from each group to form a new group. *Cluster sampling* implies classifying a number of groups within a population, then selecting some of them randomly (Beins & McCarthy, 2012). *Stage sampling* includes sampling within a sampled group (Blaxter, 2006).

A non-probability sampling is used in research works where it is impractical to get a probability sample. In non-probability sampling, the probability of selection from the population is unknown (McMillan, 1996). As mentions above, non-probability sampling compromises different methods. *Convenience sampling* is a non-random technique that includes any convenient subject for the researcher to be studied. It is termed also *accidental* or *haphazard* sample. *Quota sampling* is the technique by which subjects chosen from subgroups are selected as samples (Beins & McCarthy, 2012). *Voluntary sampling* is used when subjects volunteer to be studied (Blaxter, 2006). *Purposive sampling* (also called purposeful, judgment,

or judgmental sampling) is used when the researcher selects, some subjects that are useful and interesting for the study (McMillan, 1996). *Snowball sampling* occurs when the researcher relies on few informants to show him other informants who are appropriate for the study, and the latter, in turn, guide him to others. *Dimensional sampling* is employed when the researcher is certain that within its sample exists subjects with specific traits that are interesting to the study (Scott & Usher, 2011).

The choice of the sample size plays a critical role in determining the trustworthiness of any study. If the sample is small, it cannot represent the population, and consequently, the results cannot be generalized (Lodico, 2006). However, Cohen et al (2007, p. 101) claim that “Generally speaking, the larger the sample the better, as this not only gives greater reliability, but also enables more sophisticated statistics to be used”. Borg and Gall (1979, as cited in Cohen et al, 2007, p. 105) advocate that,

As a general rule, sample sizes should be large where: 1-there are many variables, 2-only small differences or small relationships are expected or predicted, 3-the sample will be broken down into subgroups, 4-the sample is heterogeneous in terms of the variables under study, and 5-reliable measures of the dependent variable are unavailable.

One can conclude that within a homogeneous population, a large sample size is not needed. Generally speaking, a minimum of thirty cases per variable would be sufficient or ‘a rule of thumb’ to use (Cohen et al, 2007, p. 102).

The target population in the current research work is the second year license students of English as a foreign language at Tlemcen University. It was a homogeneous population constituted of two hundred and forty-one subjects, aged between eighteen and twenty-one years old, and almost all of them had been exposed to the same educational system since their primary school. Hence, they have the same English input. The intention of the researcher was to address the third year students as being the students the most experienced in speaking among the three levels. Unfortunately, in the year of doing this research work (2014/2015), the speaking skill was not included as a subject matter at this level, i.e. at third year level, so the focus was on the second year students as a population under question.

Being very hard to reach all the subjects, the researcher found it necessary to choose a sample from them. Kothari (2004) argues that the sample “will have the same composition and characteristics as the universe. This is the reason why random sampling is considered as the

best technique of selecting a representative sample” (p. 60). Therefore, the strategy employed to sampling was the ‘*simple random sampling*’, where thirty students were selected out of the whole population. Having used this strategy to sample from a homogeneous population was likely to generate representative data and generalizable findings.

### **2-4-3 Data Collection**

Data (the plural form of datum) is defined by Oxford dictionary (2006, p. 371) as ‘*facts or information, especially when examined and used to find out things or to make decisions*’. In the research field, Griffiee defines a data as:

The lifeblood of research. Data connects theory (ideas about the world) to practice (the world). Without data, we have to take the researcher’s word for whatever claims she is making. Data allows us to look over the researcher’s shoulder and see what he saw. Data makes research empirical, and empirical research is highly valued because it represents something outside our opinions and ourselves (2012, p. 128)

Various data collection instruments can be used. Documents, interviews, observation, and questionnaires are the main ones (Blaxter, 2006:165). Data collection instruments are any physical or non-physical means used for generating quantitative or qualitative data, which will be analysed and interpreted (Griffiee, 2012).

Marczyk et al differentiate between two types of data: Metric, and Nonmetric,

Nonmetric data (also referred to as qualitative data) are typically attributes, characteristics, or categories that describe an individual and cannot be quantified. Metric data (also referred to as quantitative data) exist in differing amounts or degrees, and they reflect relative quantity or distance. Metric data allow researchers to examine amounts and magnitudes, while nonmetric data are used predominantly as a method of describing and categorizing (2005, p. 97)

Quantitative data refer to quantity and are expressed using numbers while qualitative data refer to quality and are used in describing individuals or events.

Many researchers value mixing the two methods of collecting data (quantitative and qualitative). The advantage is that the research will benefit from the strengths of qualitative and quantitative methods. The other advantage is the flexibility that the researcher enjoys in



choosing different methods of data collection. Likely results emerged from mixing both methods are more satisfying and convincing (Lodico et al, 2006).

The researcher in the current research work used three instruments for collecting data: an observation of speaking test performances, a questionnaire addressed to second year students, and an interview with the teachers of speaking module.

### 2-4-3-1 The Observation

Observation is watching what people are doing, listen to what they are saying and, at times, ask for a clarification (Gillham, 2006). Griffiee (2012) defines observation as “the systematic, intentional, and principled looking, recording, and analysis of the results of our observation for the purpose of research” (p. 178). By using the observation method, one should consider the following points: what the emphasis of the observation will be, how to record information (e.g. note taking, videotaping), how much time is needed to do it and how data from this observation will be utilized (Duff, 2008).

Griffiee classifies Observation into two categories: *participant* and *non-participant* observation, “a participant observer takes an active role in the activities whereas the nonparticipant sits in the back of the room and takes notes” (2012, p. 181). These two kinds are means of gathering data. The first step of any case study starts with a participant observation where the researcher should be very close to that case in order to wonder about any phenomenon (Griffiee, 2012). According to Gillham,

In a case study they constitute just part of the data collecting techniques. But low-key participant observation! is the one you use first: the getting-to-know phase of the research. Structured observation comes later, when you have the research issues well in focus (2000, p. 46)

Yet, Kothari (2004) makes a distinction between two other types of observation: *structured* and *unstructured*. In a structured observation, the researcher determines the criteria that will be observed and the way they will be observed. In an unstructured observation, the researcher observes the facts without any pre-planned criteria.

When using the observation method, the subjectivity of people is often avoided since the researcher him or herself records, not told about, the events. In addition, this method is more relevant for people who do not want to speak (Kothari, 2004). Nevertheless, observation has

also a number of limitations. The presence of an observer affects the discussion or the interaction in a classroom (Duff, 2008). Data provided from it are narrow. At times, unexpected elements may intervene during observation and make troubles (Kothari, 2004). Observation offers only superficial data (Richard, 1998, as cited in Griffiee, 2012)

In the present research work, the researcher used a participant observation as being a student at Tlemcen University, and a structured, non-participant observation as a tool to get a close insight on what, exactly, is happening in speaking tests. The observation objective was to see the forms of tests used, their content, the teachers' behaviour, the performance of the students, and their psychological state when performing such tests. The researcher had the opportunity to observe five speaking tests administered to the second year students with five different teachers. All of these tests were completed at the end of the second semester. The technique used to record the observation was note taking.

#### **2-4-3-2 The Students' Questionnaire**

A questionnaire is a means by which the researcher collects data from the respondent indirectly. It is a set of questions posed by the researcher and answered by the respondent (Brace, 2004). Gay and Airasian (2000, as cited in Griffiee, 2012) define the questionnaire as "several questions related to a research topic" (p. 135)

A questionnaire can take one of the three types: structured, semi-structured, or unstructured. An unstructured questionnaire is regarded to as "an open invitation to write what one wants" (Cohen et al, 2007, p. 321). A structured one is the inverse since no choice is offered to the respondent to use his or her own way to answer; answers are listed and the respondent has to tick one or more of them. A semi-structured questionnaire is a case in between. This classification is related to the kind of questions used. Several types of questions may be used, including: *closed* questions, *open-ended* questions.

Closed questions refer to the questions that recommend a set of responses the respondent has to choose only from it. *Dichotomous*, *rating scale* and *multiple-choice* questions are examples of closed questions. Using closed questions in a questionnaire has several advantages. They are helpful in terms of quickness they provide to count the frequencies of responses and to analyse them. Open-ended questions, on the other hand, offer respondents freedom to answer, as they like. They are useful in case where no possible response is available (Cohen et al, 2007).

The current research work's questionnaire aimed at getting an in-depth insight about how student view the speaking skill, the way it is taught and tested, and the difficulties they face

when performing speaking tests. It was addressed to thirty out of two hundred and forty one second year students belonging to the English department at Tlemcen University. It comprised sixteen questions. Ten of them were multiple choice, five close-ended, and one open-ended.

The students' questionnaire was divided into three parts; the first part contained four questions dealing with the students' attitude toward the speaking skill. The first question was about the reason why the students chose to learn English. The second sought their viewpoint about the meaning of *learning a language*. The third question checked out their preferable skill among the four skills of learning a language. Lastly, the fourth one, sought whether they were satisfied with their level in performing the speaking skill.

The second part included six questions concerning the teaching and learning of the speaking skill. The first question in this part (which is the fifth question in the whole questionnaire) tried to know the students' view on the way the speaking skill is taught. The second one attempted to highlight the more preferable oral task to the students. As for the third through the fifth question, the focus was on the students' personal efforts to improve their speaking performance level. The last question sought the students' opinions on which component is more important: fluency or accuracy

As for the last type of questions, the main interest was the speaking skill tests. It was composed of six questions. The first question (which was the question number eleven in the questionnaire) asked whether students liked speaking tests or not. The second sought their feelings when performing a speaking test. The third question searched for the difficulties the students encountered in speaking performance. The fourth one asked the students about the method they preferred more in speaking tests. The fifth question wanted to highlight the students' satisfaction with the way the teachers marked their speaking performances, and the last question was an open-ended question that required the students to mention to which extent speaking tests can motivate them to learn more this skill. The reason behind using a majority of closed questions was to get more focused responses.

### **2-4-3-3 The Teachers' Interview**

The interview is a sort of purposeful conversation (Duff, 2008). Nunan describes an interview as "the elicitation of data by one person from another through person-to-person encounters" (1992, p. 231). Griffiee (2012) introduces two broad categories of interviews: *standard* and *non-standard*. Standard interviews include *structured* and *semi-structured*

interviews. A structured interview involves questions that are prepared and organized in advance. On the other hand, a semi-structured interview includes prearranged questions; the interviewer is allowed to ask for clarification or even introduce other questions. Non-standard interviews are not structured. They include interviewing many persons at the same time. For Nunan (1992), interviews are classified relating to their degree of formality; three types are presented: *structured*, *semi-structured* and *unstructured* interviews. Educational researchers often use the standard semi-structured interview as it encompasses both features: predetermined questions and the ability to make changes to get more clarification (Griffiee, 2012).

The use of an interview has many benefits. Interviews can collect data that are not easily reached using other methods such as a questionnaire or observation (Blaxter, 2006). In addition, the interviewer has the opportunity to make the questions clearer (Duff, 2008). Nevertheless, the interviews have also some limitations. The first one is the difficulty of interviewing and taking notes at the same time. Moreover, the limitations may occur during the conversation such as interruption and time constraint (Duff, 2008).

The current research work follows the technique of the semi-structured interview, where the researcher had more flexibility to add some questions relying on the interviewee responses. Therefore, an atmosphere of ease and comfort filled in the interview. Eight teachers of the speaking module were addressed. The major goal was to investigate how the speaking skill is assessed while the partial goals were, generally, to seek the major difficulties encountered when teaching and testing speaking using the materials available at the university. Each interview lasted about twenty minutes. As a technique to keep the complete conversation, the researcher used a phone recorder, after getting the consent of seven teachers, and a note taking method with one teacher.

The first question looked for how experienced the teachers of speaking were. The second one tried to explore the tasks usually used in teaching speaking. The third one searched for the difficulties found by teachers in teaching speaking. As for questions four to seven, the focus of attention was on the speaking tests. The eighth question sought the teachers' views on which type of assessment is more effective formal or informal, and the ninth one asked them about how to perform reliable and valid tests.

**2-5 Conclusion**

Assessing speaking is a very delicate task that requires to be dealt with using a precise scientific way. Pegged to this, an empirical study had better to be embarked to bring to light what is actually happening, and what should be done in fact. In this vein, this chapter has shed the light on the educational situation in Algeria, with specific reference to Tlemcen University, then described the participants under investigation, and finally explained the different instruments used to collect the necessary data to such an issue. These data will be analysed and interpreted in the next chapter.

# CHAPTER THREE

## DATA ANALYSIS AND INTERPRETATION

## **CHAPTER THREE**

### **DATA ANALYSIS AND INTERPRETATION**

<b>3.1 INTRODUCTION.....</b>	<b>54</b>
<b>3.2 DATA ANALYSIS.....</b>	<b>54</b>
3.2.1 THE QUESTIONNAIRE ANALYSIS.....	54
3.2.2 TEACHERS' INTERVIEW ANALYSIS.....	65
3.2.3 TEST OBSERVATION ANALYSIS.....	69
<b>3.3 DATA INTERPRETATION.....</b>	<b>73</b>
<b>3.4 CONCLUSION.....</b>	<b>77</b>

## Chapter Three: Data Analysis and Interpretation

### 3-1 Introduction

This chapter is allocated to analyse and interpret the data collected from the different instruments used in this research work. In the analysis section, the researcher analyses each data type separately using tables and graphics that help displaying the results in a clearest way. Then, in the interpretation section, she interprets them altogether relying on correlation and comparison between them in order to infer the facts. The facts inferred from this study will provide the answers to the researcher's questions raised in advance in this research work.

### 3-2 Data Analysis

#### 3-2-1 The Questionnaire Analysis

The current questionnaire was answered by thirty students who were chosen randomly from different groups. Concerning the first part of the questionnaire dealing with the students' attitude toward the speaking skill, the learners' answers were varied. The first question, which asked the students to give their reasons for choosing English to learn at the university, was answered as follows: 43.33 % of them chose to learn English to be English teachers, 36.66 % of them considered English as a global language that should be learned. However, 16.66 % of them chose English because they plan to go abroad after finishing their study, and 13.33 % had other reasons. The following table summarizes the above results:

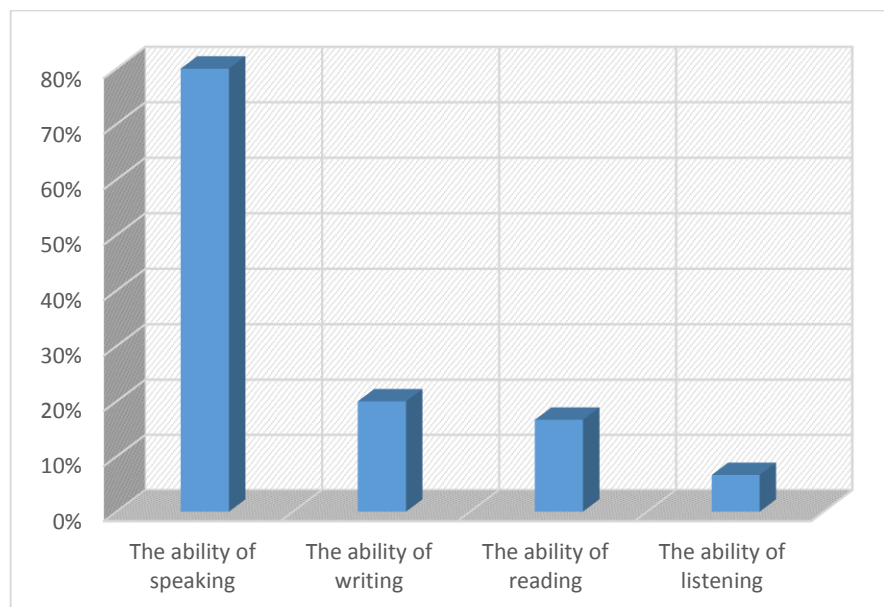
Students' reasons of learning English	Number of students	Percentage
To be an English teacher	13	43.33 %
English is the global language	11	36.66 %
To go abroad	05	16.66 %
Other	04	13.33 %

*Table 3.1: Reasons for learning English*



It is worth noting that some respondents ticked more than one answer. One can observe that more than a half of the respondents do not intend, by learning English, to be an English teacher in the future.

The second question, which asked the students to give their opinions about the meaning of learning a language, twenty-four (24) students, that is 80 % of the whole number, considered that learning a language means to be able to speak in that language. Six (06) of them, that is a percentage of 20 %, thought that it is the ability of writing which reflects the meaning of learning a language. Five (05) students, representing 16.66 % of the total number, claimed that learning a language means to be able to read in that language, and two (02) others representing 6.66 % believed that the ability of listening is the relevant answer. It is worth noting that some respondents ticked more than one answer. The following bar graph clarifies these results:

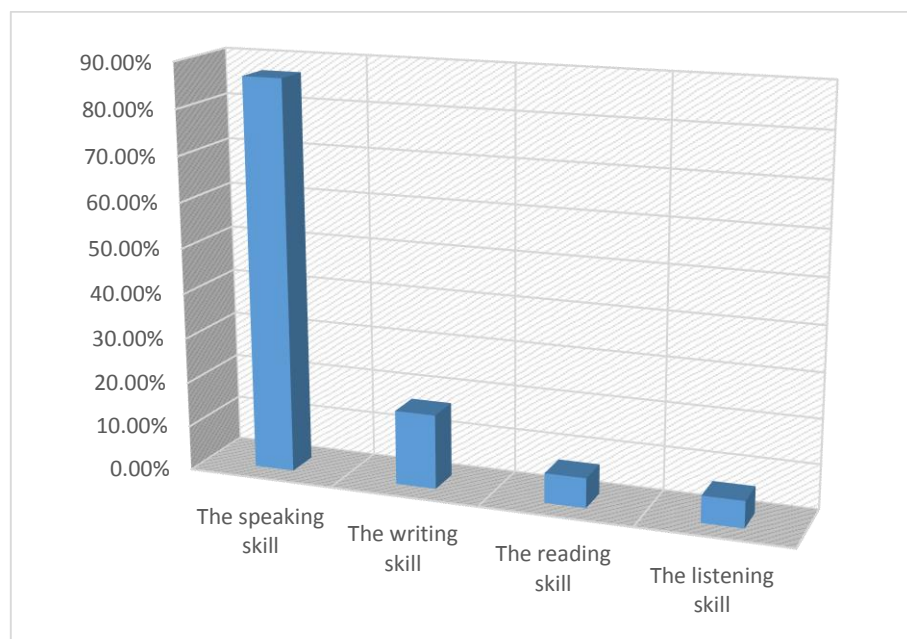


**Bar graph 3.1: The meaning of learning a language**

What is remarkable is that the vast majority of the students consider the mastery of productive skills as a sign of the mastery of a language.

As for the third question, concerning the skill the more preferred to master by the students, twenty-six (26) students representing 86.66 % of the total number preferred to master the speaking skill because, as the majority of them mentioned, it is the means of communication with others, namely with native speakers. Others stated that they wanted to improve their level of speaking. Few of them avowed that they want to be native speakers for they plan to go abroad. Few others believed that language is spoken not written.

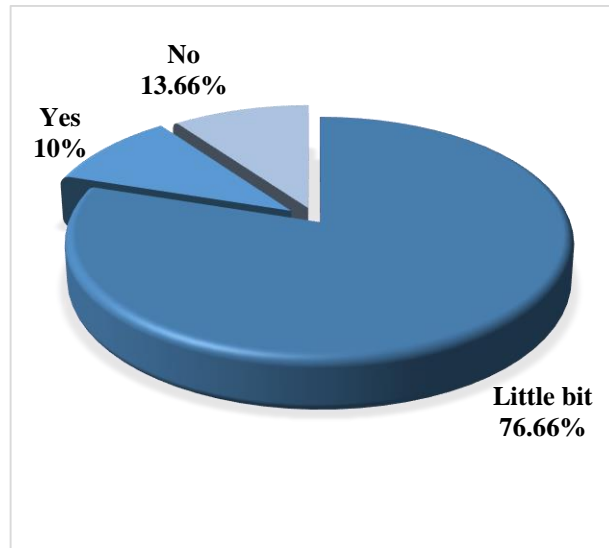
The choice of the writing skill came at the second position with a big difference in percentage comparing with the choice of the speaking skill. Only five (05) students representing a percentage of 16, 66 % wanted to master more the writing skill as they liked it and wanted to improve their writing level. However, two (02) students chose the reading skill because they wanted to improve their reading skill. Surprisingly, only two (02) students chose the listening skill even though it is very important in interactional situations especially with native speakers. The bar graph below mentions the found results:



**Bar graph 3.2: The skills preferable to master more by the students**

It is remarkable that the speaking skill is the most needed skill since 26 students (86.66%) wanted to master it while the second classified skill, which is the writing skill, was chosen only by 05 of them (that is 16.66%).

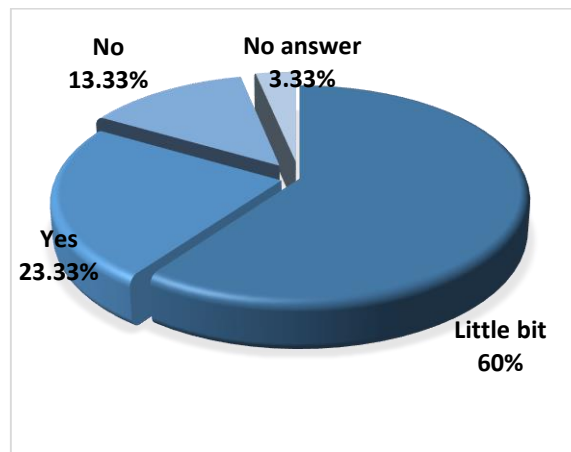
Concerning the fourth question, 23 students representing the percentage of 76.66 % of the whole number were a little bit satisfied with their level in speaking while 03 of them representing 10 % were satisfied and 04 others (13.66 %) were not. The pie chart below displays the results of the fourth and the fifth questions:



*Pie chart 3.1: Students' Satisfaction with their levels of speaking*

As it is clearly displayed, few learners were satisfied with their level of speaking.

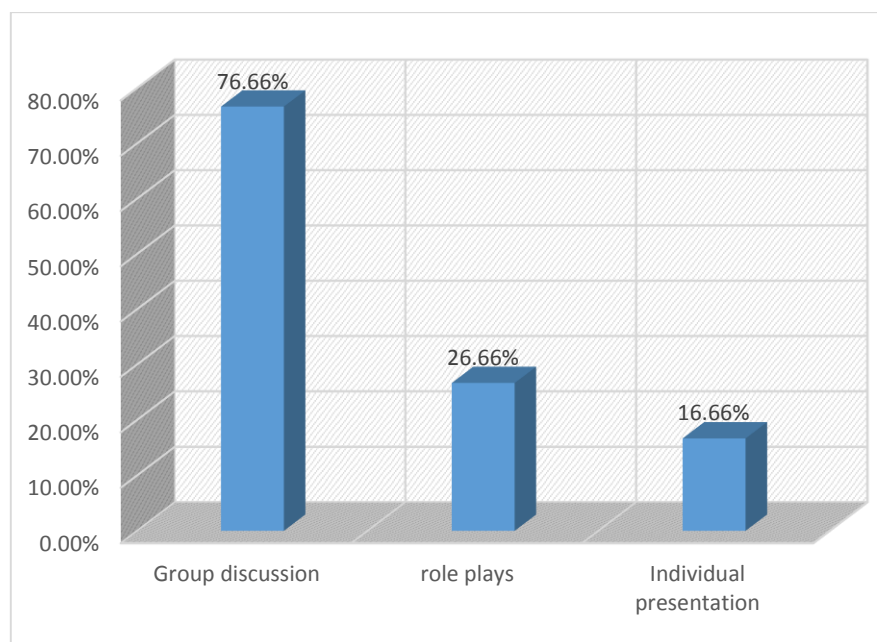
The answers to the fifth question seeking whether the students were satisfied or not with the way the speaking skill was taught, 18 students (60 %) were a little bit satisfied; 07 students (23.33 %) were satisfied, and 04 of them (13.33%) were not, while 01 student (3.33 %) did not answer. The pie chart below shows the found results:



*Pie chart 3.2: Students' satisfaction with the way of teaching speaking*

Hence, the majority of the students are in a state between satisfied, and not satisfied.

As regards the sixth question asking about which task the students prefer more to do in their ordinary speaking lectures, 23 students (76.66 %), when answering the sixth question, preferred group discussion as a task for learning speaking, 08 of them (26.66 %) liked better role-plays, and 05 (16.66%) others preferred individual presentations, however, no other suggestion was provided. The following bar graph shows these results:



*Bar graph 3.3: Preferred Tasks in the Classroom*

It is remarkable that the majority of the students preferred group discussion within their ordinary classrooms, while few of them were in favour of individual presentation.

Concerning the seventh enquiry dealing with the language used by the students to speak with their colleagues. The following table summarizes the results:

	Number of students	Speaking with colleagues (%)
Using English	9	30
Using French	6	20
Using their Dialect	20	66.66

**Table 3.2: The language used by the students to speak with their colleagues**

As it is displayed, 20 students (66.66 %) use their dialect, 09 students (30 %) of them use English, and 06 others (20 %) use French.

However, when they address their teachers (the eighth question), 19 students 63.33 % of them use English, 10 of them (33.33 %) use their dialect, and 07 of them (23.33 %) use French. The table below shows these results:

	Number of students	Speaking with teachers (%)
Using English	19	63.33
Using French	7	23.33
Using their Dialect	10	33.33

**Table 3.3: The language used by the students to speak with their teachers**

Making a simple comparison between the two tables above, one can observe somehow reversible results between the use of English and the use of the dialect; almost 20 students used English to speak with their teachers and 10 use their dialect. When speaking with colleagues,

almost 20 students used their dialect to speak with their teachers and 10 used English. However, 06 or 07 students used French to speak either with their colleagues or with their teachers.

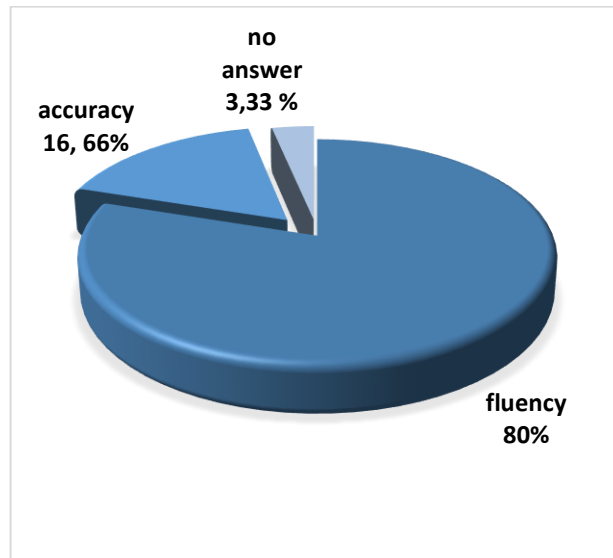
As regards speaking with native speakers, 08 students representing 26.66 % of the total number affirmed that they do speak with the natives, however, 12 students representing 36.66% did not do, and the 10 remaining students did it sometimes. The following table mentions these results:

	Number of students	Speaking with native speakers
Yes	8	26.66 %
No	12	36.66 %
Sometimes	10	33.33 %

**Table 3.4: Speaking with native speakers**

It is remarkable that the smallest number of students (8) do practice the language with natives; however, twelve (12) among them do not practice at all.

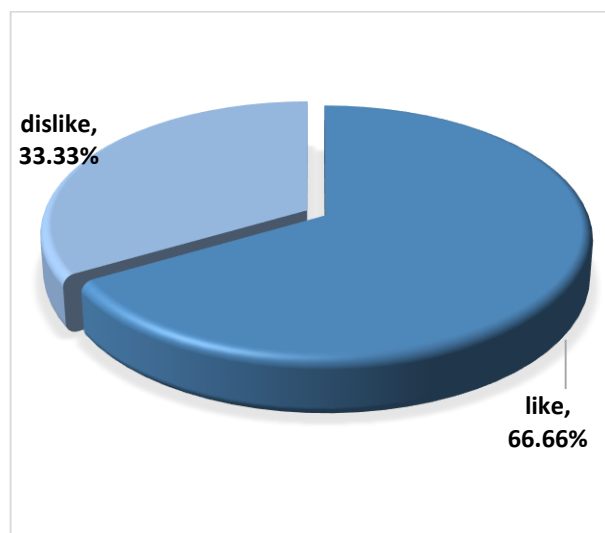
As to the tenth question, students were required to show their opinions over what is more important in a speech: accuracy or fluency. Twenty-four (24) students representing 80% of the total number agreed that fluency is more important, five (05) of them representing the percentage of 16, 66 % considered accuracy as more important, while one (01) student (3.33%) did not answer this question. The pie chart below mentions the difference between the two attitudes:



*Pie chart 3.3: Fluency vs. Accuracy*

It is clear that the great majority of the students chose fluency rather than accuracy.

As for the last part of questions, which focused on the speaking skill tests, the results were as follows: Vis-à-vis the eleventh question, the following pie chart shows the results found:



*Pie chart 3.4: Attitudes towards Speaking Tests*

A percentage of 66.66 %, that is 20 students confirmed that they like speaking tests and 33.33%, that is 10 of them affirmed that they do not. The respondents who had positive attitude said that speaking tests helped them know their weaknesses and improve their levels while the others stated that speaking tests were anxiety provoking and stressful.

For the twelfth question, the half of the students (that is 15 students) said that they feel comfortable when performing a speaking test, and the others feel anxious. The subsequent bar graph makes these results clear:



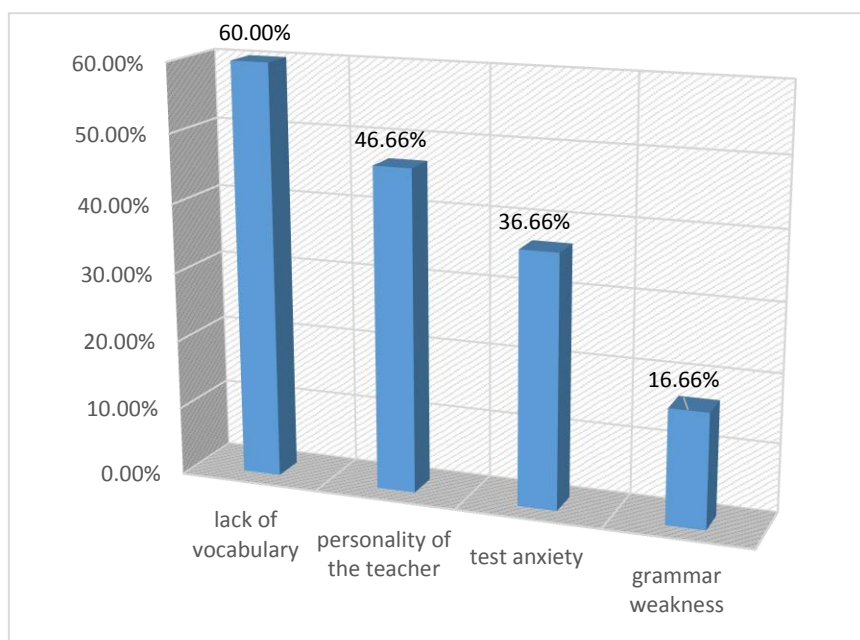
*Pie Chart 3.5: Feelings during Speaking Tests*

As it is seen, learners, in terms of their feelings, are divided into two equal parts; fifteen of them pass their tests in comfort, and fifteen in anxiety.

As to the factors affecting the students' performances in speaking tests (posed in the question number thirteen), the students' answers uncover that 18 students (60%) suffer from the lack of vocabulary; the personality of the teacher was classified as the second disturbing problem for 14 students(46.66 %). Anxiety was the problem of 11students (36.66%), and lastly grammar weakness was disturbing only 05 students (16.66%). It is worth noting that some students ticked more than one problem.

The following bar graph encompasses these results:





**Bar graph 3.4: Factors affecting the students' speaking performance**

It should be noted here that some students selected more than one factor. What draws the attention here is that a large percentage of students suffer from lack of vocabulary.

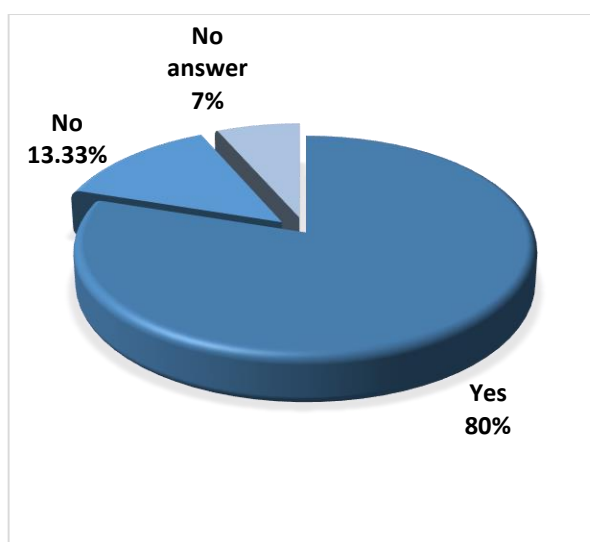
As for the question fourteen, twelve students (12) representing 40 % of the total number preferred group interview method of speaking tests. Eleven (11) of them representing the percentage of 36.66 % liked better a paired interview method. However, eight students (08) standing for the percentage of 26.66 % were in favour of an individual face-to-face interview, i.e. the teacher is the interlocutor. The table below displays these results:

	Number of the students	Percentage (%)
Group interview	12	40
Paired interview	11	36.66
individual interview	8	26.66

**Table 3.5: Preferred way of a test interview**

One can infer from the above table that the students preferred the collective performances than the individual ones. We notice that only eight students preferred an individual interview against twenty-three who favoured either paired or group interviews.

Regarding the fifteenth question asking whether the students were satisfied with the rating methods used by their teachers, eighty percent (80 %) standing for twenty-four (24) students said 'yes', four (04) students representing 13.33% said 'no', and two of them (02) representing 6.66 % did not answer. The results are displayed in the following pie chart:



*Pie chart 3.6: Students' satisfaction with the teachers' ways of rating*

As it can be seen, the majority of the students were satisfied with the way teachers rate their test performances.

As for the last question seeking the opinions of the students about the relationship between speaking tests and the students' motivation to learn the speaking skill, the half of them (15) did not answer; one student (01) said that speaking tests have nothing to do with their motivation because it is a matter of will. However, fourteen students (14) affirmed that speaking tests have a role in enhancing the learning of this skill. The majority of them (i.e. the 14 last students) confirmed that speaking tests motivate them to learn more and improve their speaking skill; some of them stated that speaking tests help them to overcome their stress and anxiety, and few of them said that they assist them improving their fluency as well as their vocabulary range.

### 3-2-2 Teachers' Interview Analysis

The interview with teachers of oral expression was a rich source of data. Teachers spoke about their everyday challenges, their efforts, and their wishes. The first question concerning their period of teaching oral expression uncovers that they are not experienced teachers. The older one among them has eight years of experience in that module, three others have between six and seven years, and the others have less than three years.

Concerning the tasks usually included in teaching speaking, the teachers talked about various tasks. A common task that all the teachers use is the individual or paired presentation. The students are required to perform one or two presentations during the instructional year. Other tasks such as group discussion about any topic, games, individual talk, and practicing language functions such as introducing one's self, greeting, agreeing, etc. are covered in different proportions. Some teachers use other less common but fruitful tasks. An example of this is the '*book report*'. To do it, students are required to read a book and to summarize it in a spoken report. The '*game of chair*' is another task where a student sits on a chair in front of his or her classmates and has to answer all their questions even if they are embarrassing or difficult. Another rare task is to answer a written questionnaire about any topic, and then the class is engaged in a debate on this topic. The teacher who uses this last task said that it is a motivating task and students through answering the questionnaire are given the time and the opportunity to get pre-arranged ideas that help them a lot to speak with confidence. Few teachers stress the ability to speak in front of an audience. They said that facing an audience and speaking in any way, but with confidence is their major goal and is more important than focusing only on the quality of speaking. A seldom task which is used only by a one teacher is '*idiomatic expressions*'. Using this task requires the students to learn by heart an idiom every session and use it in different sentences. This teacher views that idioms are very important in enhancing one's level of speaking because they are very used in native speakers' interactions and they are so expressive and sometimes so funny to motivate students to speak using them.

As to the test tasks, the majority of the teachers said that they perform interviews concerning only what students have seen during the course such as their own or their peers' presentations, the language functions that they practiced in the classroom or free topics of their own choice. A face-to-face interview might be the only task that teachers do in testing speaking.

The following table displays the tasks used by the teachers in teaching and testing:

Tasks included in teaching	Tasks included in testing
<ul style="list-style-type: none"> <li>• Presentation</li> <li>• Group discussion</li> <li>• Games (e.g. the game of chair)</li> <li>• Individual talk (e.g. a speech)</li> <li>• Book report</li> <li>• The questionnaire</li> <li>• Idioms</li> </ul>	<ul style="list-style-type: none"> <li>• Face-to-face interviews</li> <li>• Speaking in front of a tape recorder</li> </ul>

**Table 3.6: Tasks included in teaching and testing**

Regarding the difficulties that the teachers are facing in their teaching, the majority of them complained about the crowded classrooms. It is very difficult to manage a classroom with more than forty students, therefore, it is more difficult for the teachers to diagnose the students' needs and weaknesses and cure them. From the students' side the big number prevents many of them from getting the chance to speak. Another difficulty is the time allotted to this module; it was only three hours per week divided between listening and speaking. The teachers cannot do their job effectively within this short time. The third common difficulty is the lack of materials devoted to teaching such a module. There are two laboratories equipped with technical aids, but they are not sufficient comparing with the number of the students.

There are also other kinds of difficulties related to the students themselves. Some of them are not motivated to learn and to make efforts, others lack confidence and are afraid of making mistakes so they keep silent and prefer only to listen, and there is the kind of learners who want to speak but they lack the language competence. The weak linguistic competence puzzles the teachers; they do not know whether to spend the time of the session in correcting their students' errors or going on without paying attention to this matter. Another problem, which was raised by a single teacher, was the heterogeneity of the learners what makes meeting the needs of everyone a hard task.

As for the difficulties found in testing speaking, the teachers discussed numerous ones. The main difficulty that all the teachers confront is time; speaking tests are time-consuming. Testing more than forty students one after one takes much time. A second difficulty is the lack of materials, namely ICTs. Another common and very important problem is ‘*what to assess and how to assess it*’. Teachers have not a collective agreement about which ground they should rely on to score their students speaking performance. Every teacher has his or her own criteria to do that. Students’ anxiety is also a common problem. Some students are at ease when speaking with their teachers in their ordinary classes, but when it comes to the exam, they become afraid, under stress, and they end up losing their tongues.

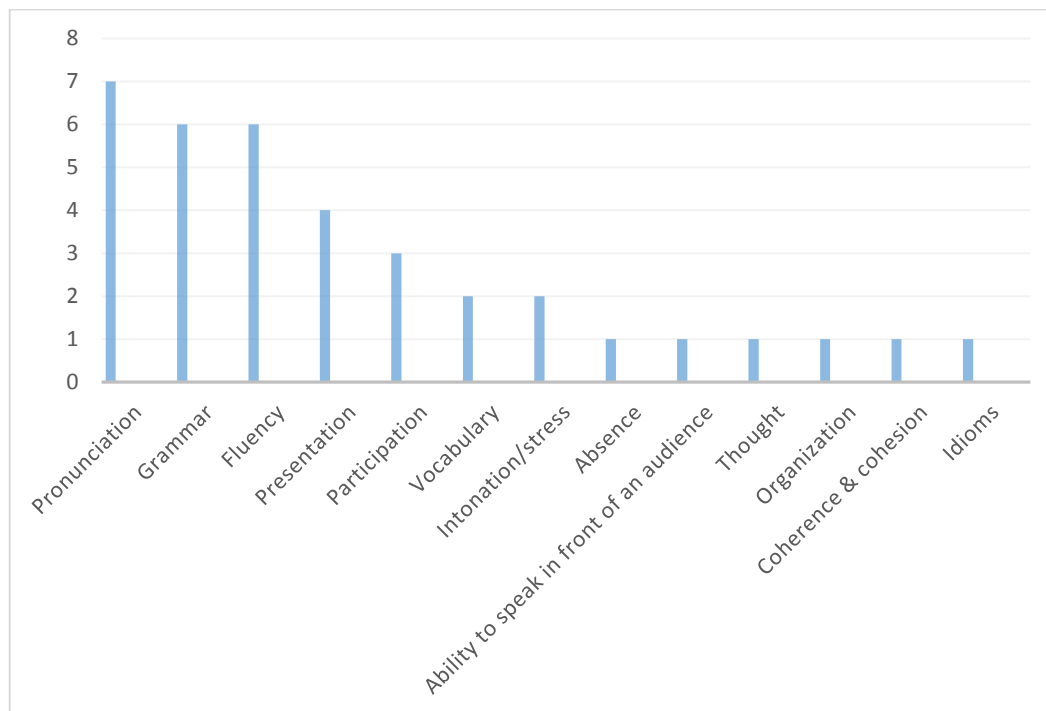
The following table displays and compares between the difficulties found by the teachers either in teaching or in testing speaking:

Difficulties in teaching speaking	Difficulties in testing speaking
<ul style="list-style-type: none"> <li>• Crowded classrooms</li> <li>• Lack of materials</li> <li>• Short time allotted per week</li> <li>• Lack of motivation / Weak language competence</li> <li>• Students’ silence/fear of mistakes</li> </ul>	<ul style="list-style-type: none"> <li>• Crowded classrooms</li> <li>• Lack of materials</li> <li>• Time-consuming</li> <li>• What to assess</li> <li>• How to assess</li> <li>• Students’ anxiety</li> </ul>

**Table 3.7: difficulties in teaching and testing speaking**

Since teaching and testing are inseparable, one can observe that the difficulties are almost similar in both cases, except some differences related to the specificity of each one.

As regards the rating criteria used to assess the speaking ability of the students, each interviewed teacher, except one, has his or her own one. The following bar graph clarifies the criteria used by the teachers in their rating scales:



**Bar graph 3.5: Criteria used to assess the students' speaking ability**

As it is seen, pronunciation is an important criterion in the estimation of each one of the seven teachers. Grammar and fluency are also important; six teachers evaluate them. Four teachers introduce presentation, which is a classroom task, as a criterion on their rating scale. However, only two teachers include vocabulary and intonation/stress. Other less common criteria are seldom used such as absence, ability to speak in front of an audience, coherence and cohesion, thought, idioms, and organization.

One among the eight interviewed teachers avowed that she does not use a rating scale at all, but she analyses the performance of each student and score it holistically. Another teacher said that he does use the rating scale, but he focuses on the attendance of the student at the ordinary classes more than his or her speaking fluency because it could come about that a good speaker, who has never attended an oral expression session, gains the highest mark in the exam; this is unfair.

Regarding the enquiry about which is more important in a speaking performance: fluency or accuracy, the majority of the teachers said that both of them are important, but it is not logical

to focus, for beginners, on accuracy. Some of them focus only on fluency, and only one teacher said that he stresses accuracy on the account of fluency.

As to the comparison between formal and informal speaking tests, the majority of teachers prefer doing informal test. One teacher among this majority said, “*Informal tests are better and fruitful, but formal tests are imposed on us*”. Few of them said that both types are important and could be used depending on the students’ needs. A teacher said that informal tests fit more the students who are anxious because they are more at ease outside formal tests.

Concerning how to reach more valid and reliable tests, the majority of the teachers failed to grasp what does this mean. Many of them, attempting to answer, asserted that assessing such a skill should be done over the entire instructional year far from the threatening conditions of formal tests. Few of them stated that grades are never objective and they were trying to find solutions for this issue. Novice teachers acknowledged that they have no idea about these issues and that they need to train. A teacher suggested that they needed to change their old methods of testing and scoring. Another one stressed a collective agreement on test format, criteria and scoring process. However, only a single teacher emphasized the use of technology. She stated that technology helps teachers and students alike. Teachers, through using ICTs, avoid time-consuming methods and can analyse their students’ performances at ease because all things are recorded. She added that students feel more comfortable in front of a computer than in front of a teacher, and they have the chance of repeating their performance.

### **3-2-3 Observation Analysis**

I observed five speaking tests administered at the end of the second semester. In the first observation, the teacher told me that he had informed his students about the topics included in the test to prepare relevant ideas. The test was a face-to-face interview where the teacher did not take an interactional part, but he took on the role of reading the question chosen by the student, asking for clarification from time to time, or changing the question in some cases. The topics were numbered from one to ten, and the students had to choose a number randomly, listen to the question read by the teacher, and then answer after a few minutes of thinking. The teacher was trying to create a friendly testing atmosphere to make his students comfortable.

I observed the performance of nine (09) students. What drew my attention is that although the teacher had prepared ten topics as a test content, the focus was mainly on two topics: four

students chose the topic number (01) and four others chose the topic number (06), and only one of them chose another topic. Regarding the linguistic competence, the majority of students had a lack of vocabularies, and a weak fluency, while they have an acceptable pronunciation. From a psychological point of view, four students out of nine were anxious; however, the others were at ease. The teacher, in order to evaluate the performance of each student, used symbols (- / +) in front of each criterion in the checklist that involved four criteria: fluency, pronunciation, vocabulary, and grammar. Then he analysed the overall performance at home and calculated the final grade for each one.

In the second observation, the teacher interviewed either one or two students at once. I had the opportunity to watch six interviews managed with nine students. The teacher's firmness was dominating the test's atmosphere. The interviewees were asked to speak about the same topic, which was somehow difficult, but had been seen in their class; with slight changes from a student to another. Two main remarks attracted me in this testing: (1) the majority of the students' performances seemed to be learned by heart, and (2) seven out of the nine students were anxious. Linguistically speaking, many of them had a lack of vocabulary and a low level of fluency, but they had a good pronunciation. To score the performances, the teacher used a rating scale that included five criteria: presentation, intonation, grammar, pronunciation, and fluency. The fluency criterion was assigned the lion's share of points: eight points, however, the other four criteria were assigned three points for each.

The third observed testing session was different. It looked like an ordinary speaking lesson. The teacher interviewed four students at a time about any topic that had not been prepared for previously. He used a funny way to motivate his students to speak. He allowed them to speak about their classroom previous presentations or about any topic they find it easy. Psychologically, the majority of students were at ease. Linguistically, they had a lack of vocabularies, an acceptable pronunciation. As to fluency, it was not easy for the observer to know their level because the students took short turns of speaking in which they uttered a few words while the teacher took the most speaking time.

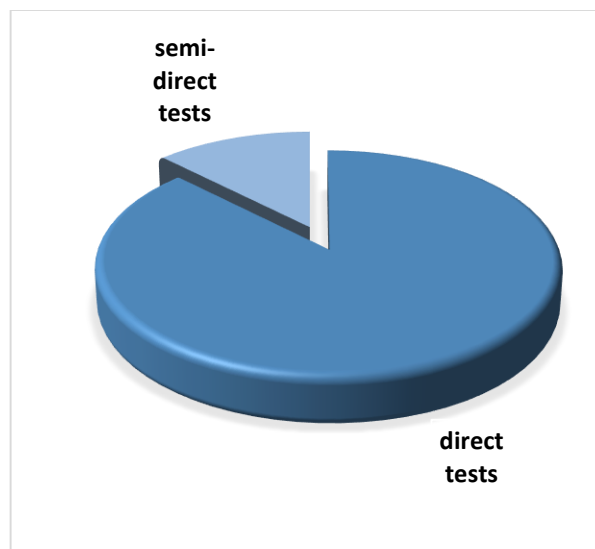
The fourth observed test was a face-to-face interview in which the interviewer is the teacher-examiner. I observed the performances of eight students. The teacher was a type of a serious and an easy-going person at the same time. She asked each student two questions concerning varied topics such as their educational experience, social life, their interests and their presentation topics. The majority of the students was answering the teacher's questions at ease, up to the point that one of them told the teacher about her private life problems and her wishes



without any reservation. . However, few of them were anxious. In terms of language, many of the students used a usual range of vocabulary. Their pronunciation and fluency varied from acceptable to good.

The fifth observation took place in the laboratory. The teacher, not as usual, used a semi-direct test. Students took their seats, prepared their technical aids such as the headsets, the speakers, and the recorders; listened to the teacher’s question, then started speaking quietly each one to his own tape recorder. They had to record their performances within no more than two minutes, and they had the right to listen to themselves and repeat again if they do not like the first attempt. The majority of them looked comfortable; however, their performances were not observable due to the nature of the test.

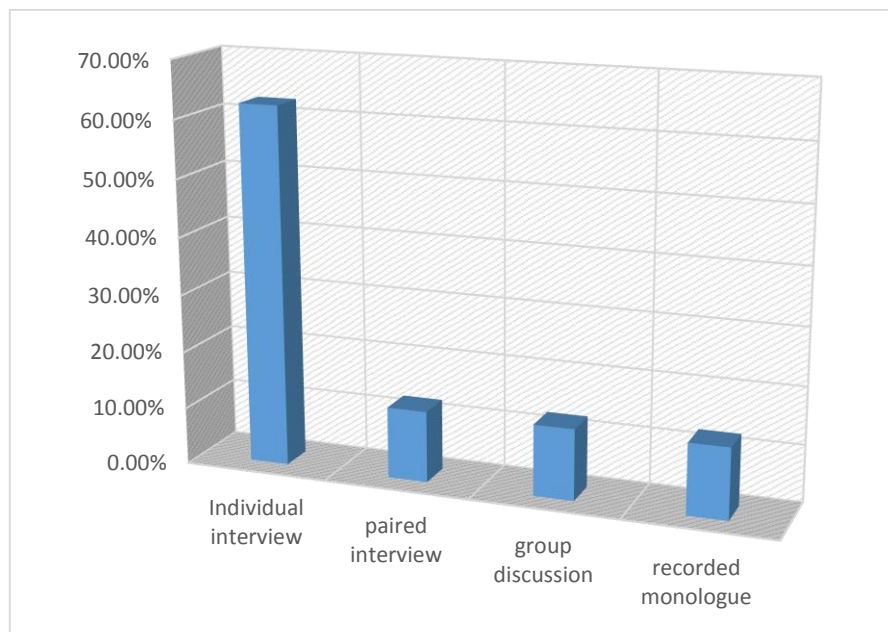
As a conclusion, the tests observation helped the researcher to collect many evidences. Firstly, in terms of the test type used, seven out of eight teachers use the direct speaking tests. The teachers performed a one-way interview with their students. A one-way interview means that the interaction is little, and is directed only in one way: from teacher examiner to the examinee. The examiner does not interact with the student; his or her role is restricted in asking questions, requesting for clarification, or making remedies in some cases as changing the question or adding some words that help the student to speak more. A single teacher use a semi-direct type, which includes ICTs to facilitate the teacher’s job. The following bar graph clarifies this result:



*Pie chart 3.7: Types of tests the teachers use*

The teachers stated that the lack of the necessary materials and rooms, and the time wasted in the arrangements of ICTs were the reason behind choosing direct tests type.

Secondly, concerning the tasks included in the tests, the majority of the teachers used individual interviews; one teacher uses paired interview, and another uses group interview. However, the teacher who uses the semi-direct test, use recorded monologue as a task. The following bar graph summarizes this result:



**Bar graph 3.6: Test tasks**

Five out of eight teachers, that is 62.50%, use individual interviews because they believe that this task type helps them to evaluate their students' performance more effectively than paired or group discussion. The teacher who uses paired interview do that because the topic of the test is that of the presentation performed in pairs during the ordinary classes. As to the teacher who prefers group discussion, he is convinced that testing should be a learning experience and that the real assessment should be done in the course of the teaching sessions. As to the range of vocabulary used, the researcher observed that the majority of the students have a lack of vocabulary.

As far as rating criteria were considered, the majority of the teachers stressed pronunciation, grammar, presentation, and fluency. What was more remarkable is that each teacher has his or

her own rating criteria that involves the criteria listed above and others. Only two teachers out of the eight observed use the same rating criteria. However, one teacher rate the students' performance in a holistic way. The results from this were the same as those from the teachers' interview.

### **3-3 Data Interpretation**

Relying on the results obtained from the research instruments (the questionnaire addressed to the student, the test observation, and the interview with the teachers of the module), the researcher, in this section, will examine the credibility of the hypotheses that was introduced in the beginning of this research work.

Concerning the first hypothesis that proposes that teachers assess the speaking performance of their students as a productive skill and not as an interactive skill, the results extracted from the test observation and from the teachers' interviews were largely confirmatory. Teachers focused mainly on speech production to test their students' speaking performance. The evidence for that is that the rating criteria the most used by the majority, which is 75% of the teachers, are pronunciation, grammar, and fluency. These criteria are indicators of the linguistic competence of the students; however, the speaking skill includes also the communicative competence.

From another perspective, the students' answers to the questions (1), (2), (3), and (4), displayed the following results:

- More than 43% of the students chose to learn English because they want to be English teachers, and more than 36% chose it because it is a global language.
- The majority of the students, that is 80% of them, view that learning a language means to speak in that language.
- The majority of the students, more than 86%, prefer to master the speaking skill
- More than 76% of them are not satisfied with their speaking level.

Likewise, the students' answers to the questions (11) and (16) related to their opinions on speaking tests revealed that:

- 66.66% of them like speaking tests
- Fourteen students, among fifteen students who answered the question (16), that is 93.33% of them, stated that speaking tests represent for them an opportunity to practice this skill, to improve their levels, and to overcome anxiety.

These results are all signs illustrating that the students like strongly this skill and yearn to master it for the sake of an effective communication.

However, the results from the questions (7) and (9) revealed contradictory facts:

- More than 66% of them use their dialect to speak with their colleagues.
- Only 26.66% of them do speak with native speakers.

The majority of the students do not make efforts to reach a native-like speaker level. This can be interpreted by the fact that learners, usually, in their learning process, are more interested with what will be tested, i.e. the test content, therefore they prepare themselves to pass the module. In other words the students' learning is, often, a test-driven learning. As the present speaking tests are testing only speech production, the students are not motivated to practice the interactional side of the speaking skill. From this fact, one can infer that the current speaking tests have not a positive washback on students' learnability, that is, they do not enhance their motivation. A second interpretation for these results is that learners are not autonomous in their learning because they do not perform the activities that boost their autonomy .

As regards the second hypothesis that assumes that teachers are still using out of date tasks and rating criteria, the findings do support that. Firstly, in terms of tasks, the test task the most used by the majority of the teachers is individual face-to-face-interview. This task has its advantages, however, has many drawbacks. The first obvious inconvenient of a face-to face-interview is being time-consuming, especially when the number of examinees is huge. All the teachers, through the interview, complained about this. This was also seen during the observation; an interview with a single, normal student took at least five minutes, let alone an anxious or a shy one. Multiplying five minutes by forty or more would result in a too long time and an exhausted examiner who, by logic, did not treat the latter students the same as the formers in terms of either the length of the interview or the test scoring.

The second inconvenient for this task is that it does not reveal the interactive skill of the learners. A face-to-face, one-way interview is not useful for the sake of interaction. Viewing the students' answers to the fifth question of the questionnaire, only seven (7) students, consisting 23.33% of the total number, are sure of their satisfaction with the way the speaking skill was taught while the others, that is 76.66% are not satisfied or are confused. Their answers to the sixth question may be an explanation for this. A percentage of 76.66% of the students,

expressed their wish to perform more interactional tasks; they like better group discussion. This can be interpreted that this task has the advantage of motivating the learners to speak more, and thus enhancing their interactional skill. A less percentage of the students, that is 26, 66%, like better role-plays which is also an interactional task. This is maybe for (1) its nature as being not spontaneous, i.e. the speaking turns are learned by heart and are not uttered extemporaneously, and (2) in that it does not involve all the students. Yet, only 16.66% of them are in favour of the 'presentation' task, which is favoured by the majority of the teachers in either teaching or testing, because this type of tasks do not include much interaction. This disinclination to 'presentation task' by the students can be explained that they want more enthusiastic tasks, and that they are bored with this task type.

The third inconvenient is that this type of tasks are anxiety creating. The results from the questionnaire support this. When students were asked, through the questionnaire (question number 14), to choose the interview type the most relevant for them among the three types: individual, paired, or group interview; the majority of them chose either paired or group interview. Likely, half of the students, when answering the twelfth question in the questionnaire, stated that they feel anxious when sitting for a test. This means that they avoid being alone in front of the teacher, and that they feel better when being with their colleagues. One of the interviewed teachers said that for many students, facing the teacher examiner alone is one of the factors that arise anxiety. The researcher observed this too; many learners were anxious in front of their teachers. The choice of a face-to-face interview task is not relevant for all the students. Many interviewed teachers stressed the necessity of changing this method. They said that for some learners it is better to be assessed within their ordinary classes because they speak at ease, while they become anxious when they face alone their teachers.

Secondly, in terms of the rating criteria used to score the students' speaking performance, the majority of the teachers do not keep abreast of the recent approaches to testing. They use criteria that are not useful for assessing the '*communicative competence*' of the learner. This latter has been defined, as we have seen in chapter one, by many intellectuals as being the ability of using the linguistic competence in different communicative contexts. However, there are different approaches to test this (see approaches to testing, chapter one). Many criteria are introduced to assess the learners' communicative competence. *Grammatical competence*, *sociolinguistic competence*, *strategic competence*, and *pragmatic competence* are examples of the recent criteria included in testing the learners' communicative competence.

Besides, the majority of the students, that is 80% of them, when they were asked whether they like the way their teachers rate their performance, answered 'yes'. One of the possible interpretations for this is that the rating criteria used by the teachers are easy, and make it easy for them to pass. Supposing that teachers use the up to date criteria listed above in addition to the previous ones, the results would be largely different, and passing the module would require much effort.

As regards the third hypothesis anticipating that the major difficulties that hinder the students' speaking ability are '*anxiety*' and '*lack of vocabulary*', findings from the students' questionnaire, showed that '*lack of vocabulary*' was the most disturbing difficulty for the majority, that is 60% of the students. The difficulty that was classified next was '*the personality of the teacher*'. This difficulty was supported by 46.66% of the students. Certainly, students meant by '*the personality of the teacher*' the negative side of this personality, such as inflexibility, unfairness, frown, etc. However, '*test anxiety*' was classified, by 36.66% of the students, as the third troubling problem, and '*weakness in grammar*' as the last interesting problem by 16.66% of them.

Additionally, the learners' answers for the tenth question, that asked them to compare between fluency and accuracy, revealed that most of them, that is 80% of them, view that fluency is more important than accuracy. Fluency cannot occur without having a large range of vocabulary in mind. This is compatible; on the one hand, with the students' complain of their '*lack of vocabulary*' problem, and on the other hand, with their carelessness of '*weakness in grammar*', i.e. accuracy. However, the teachers stated that both fluency and accuracy are important in one's performance, for they are looking for a high speaking level.

As for the findings taken from the test observation, the researcher observed that the majority of the students had a '*lack of vocabulary*'; and that '*the firmness of the teacher*' was a factor that increases the students' anxiety.

The teachers' interview, alike, asserts that many students become anxious when facing the teacher in a test, which means that the personality of the teacher has a negative impact on the students' performance.

Combining the three sources of data, one can conclude that the most disturbing difficulty for the learners is first the *lack of vocabulary*, and second the *personality of the teacher*.

**3-4 Conclusion**

This chapter has provided answers to the questions aroused by the researcher concerning the assessment of the students' ability of speaking; namely what teachers assess, the method of assessment, and the major difficulties that slow down this ability. At the same time, it has confirmed that teachers are assessing only the productive skill of the students and not the communicative one. It has also asserted that the tasks and the rating criteria used by the majority of the teachers are not in line with the recent practices in testing. Lastly, this chapter has uncovered the major difficulties facing the students when being tested.

The next chapter will suggest some remedies that serve to get rid of these problems, and that help improving the assessment procedures, and thus the students' speaking ability.

# CHAPTER FOUR

## SUGGESTIONS AND RECOMMENDATIONS



# CHAPTER FOUR

## SUGGESTIONS AND RECOMMENDATIONS

<b>4-1 INTRODUCTION.....</b>	<b>78</b>
<b>4-2 INTRODUCING SPEAKING AND SPEAKING ASSESSMENT.....</b>	<b>78</b>
4-2-1 PERFORMANCE-BASED ASSESSMENT.....	83
4-2-2 SELF AND PEER-ASSESSMENT.....	84
4-2-3 PORTFOLIO ASSESSMENT.....	85
<b>4-3 RATING CRITERIA.....</b>	<b>86</b>
<b>4-4 ENSURING A POSITIVE WASHBACK ASSESSMENT.....</b>	<b>88</b>
<b>4-5 SPEAKING ASSESSMENT TASKS.....</b>	<b>90</b>
<b>4-6 LESSENING THE DIFFICULTIES ENCOUNTERED BY THE STUDENTS.....</b>	<b>94</b>
4-6-1 STUDENTS' LACK OF VOCABULARY.....	94
4-6-2 THE PERSONALITY OF THE TEACHER.....	96
4-6-3 STUDENTS' ANXIETY.....	98
<b>4-7 CONCLUSION.....</b>	<b>100</b>
<b>GENERAL CONCLUSION .....</b>	<b>101</b>
<b>BIBLIOGRAPHY .....</b>	<b>104</b>
<b>APPENDICES.....</b>	<b>113</b>
<b>APPENDIX 1.....</b>	<b>113</b>
<b>APPENDIX 2.....</b>	<b>115</b>
<b>APPENDIX 3.....</b>	<b>116</b>

## **Chapter Four: Suggestions and recommendations**

### **4-1 Introduction**

After having got an overview of the actual situation in the University of Tlemcen concerning the assessment of the speaking skill, the difficulties found by the teachers and the students alike in learning, teaching, and assessing this complicated skill, a last and important step in this research work would be providing some suggestions and recommendations to adjust the current situation. The present chapter has the aim of levelling up the speaking assessment practices in our university to be in line with that in the developed universities over the world, as well as to improve the students' speaking ability. It will begin first by shedding the light on the terms 'speaking' and 'speaking assessment' as an essential step to know what should be assessed, i.e. the criteria that should be elicited in a speaking performance, then which test types and tasks are more relevant. At the end, some solutions will be provided to lessen the difficulties found by the students either in learning or in testing.

### **4-2 Introducing Speaking and Speaking Assessment**

Providing a definition for what we are assessing is a very important step that should be looked for before making any decision. By doing so, teachers become able of using the right tasks to elicit the learners' speaking ability. If teachers do not do so, the test and the scores obtained from it are considered meaningless (Fulcher, 2003). For this reason, teachers have to meet together in order to reach a consensus about standardized speaking assessment (Restrepo et al, 2003). This consensus involves how to define the construct being assessed depending on the purpose of the assessment; and therefore, indicate the relevant criteria for this assessment and the appropriate tasks that serve revealing these criteria. Douglas (2000) mentions that the construct definition is the theoretical side of the speaker's ability, and the assessment criteria are the operational definition of the construct.

However, the construct definition is affected by the conditions surrounding the assessment. Fulcher makes this point clear saying that,

More recently there has been a tendency to admit that sometimes, when contextual factors clearly impact upon discourse and test score, it may be more appropriate to include the contextual factors in the construct definition (Chapelle, 1999a). This may be most appropriate in tests of speaking for specific purposes (Douglas, 2000), where the definition of the construct, the selection of tasks and the scoring procedure are designed to collect evidence for ability to speak in a small sub-set of contexts, such as service encounters or air traffic control (2003, p. 19)

Douglas (2000) points out that the interaction between the features of the language use situation and the construct definition is necessary. In our context, for instance, English is not the language of instruction, nor is it used socially as French is. Moreover, few opportunities are available to practice it. Teachers should make this fact into consideration when defining the construct.

Many definitions of the term ‘speaking’ have been provided since the second half of the twentieth century until now. In the 1960s, the speaking skill was introduced as, “Not only... The use of the right sounds in the right patterns of rhythm and intonation, but also the choice of words and inflections in the right order to convey the right meaning” (Mackey, 1965, as cited in Bygate, 1987, p. 5). This approach has been embraced for more than ten years until when David Wilkins called attention to the ineptness of this view in solving some existing learning problems among which, ensuring “a satisfactory transition from supervised learning in the classroom to real-life use of the skill” (Bygate, 1987, p. 5) was the more important one. Bygate means by *supervised learning* the skills acquired in the classroom about the language such as vocabulary, grammar and pronunciation that he calls them ‘*motor perceptive skills*’, and by *real-life use the skills of using the motor perceptive skills* in authentic situations. He calls these skills ‘the *interaction skills*’. Thus, speaking involves the combination of these two skills. To make this clear, a comparison with a car driver can be beneficial. It is not sufficient for the car driver to know the controls, their names, their locations within the car, and their functions, but he or she should be skilful in using them in either a straight or an uneven road in different speeds, without crashing into anything that might exist in the path (Bygate, 1987).

With this in mind, teachers have to pay attention to what they are assessing. The present research work results showed that they are assessing only the motor perceptive skills. A 2003

study similar to the current research work conducted by some teachers at the EAFIT Language Center, Medellín, Colombia, revealed that,

Most teachers have a weak knowledge of assessment; that assessment is done for summative purposes rather than formative; that assessment is done unsystematically and without much planning; and that teachers use multiple activities to assess different aspects of oral language with some attention to students' level of language proficiency. The implications of these results indicate the need for educational programs in the area of assessment and the creation of an oral assessment system (Restrepo et al., 2003, p. 63)

Those researchers stated that the spoken language assessment is no more that of testing only grammar and pronunciation, but it should investigate the students' interaction competence in real life situations. Luoma points out that "From a testing perspective, speaking is special because of its interactive nature" (Luoma, 2004, p. 170). This interactive nature of the speaking skill makes it an essential component in communication. Lightbown and Spada, (1993, as cited in Restrepo et al, 2003) assert that the recent decades witnesses a growing trend towards the communicative approach to language teaching and assessment in which the focus is on real-life communication, especially interaction, conversation, and language use. Varela & Palacios state that "In many current English language teaching methodologies communication in its fullest form has become the main aim of the teaching-learning process, and interest in oral tests derives from this growing attention to the listening and speaking skills" (2013, p. 56).

Nowadays, communication is the goal the most demanded especially by young learners who are taken with the globalization wave. For this reason, teachers of speaking have to cope with the mainstream trends. Thus, they have to reconsider their practices and regulate them in a way that guarantees the reach of their learners' needs and goals. To do this, a correlation between teaching, learning and assessment is crucial and necessary; however, there is a "generalised feeling of a divorce between learning and teaching on the one hand, and assessment on the other. The fundamental reason for this is that assessment often does not feed back into the learning and teaching process" (Harris and McCann, 1994, p. 3). As such, assessment is the only benchmark that informs both teachers and learners whether they are walking in the right path or not.

From a historical point of view, Fulcher illustrates the aim of the first speaking assessments in the United States,

Testing second language speaking has become extremely important since the Second World War. When it was realised the American service personnel were often ill equipped to communicate, that their inability to speak in a second language might be ‘a serious handicap to safety and comfort’, the push for tests of second language speaking began in earnest. From the needs of the military, the practice soon spread to colleges and universities. Today, speaking tests are available from all major test producers, for a multitude of purposes (2003, p. xi)

Even if having safety and comfort was one among the very first reasons that led people to test second language speaking, this is not far from our thought as Muslim learners of English as a foreign language. Our prophet Mohammed may the peace and blessing of Allah be upon him said *“He who has learned other people's language he will be safe from their deception”*.

Concerning the type of assessment the more relevant, a formal classroom assessment does not suit a complicated skill such as speaking. Neither the type of test, nor the tasks used are serving the aspirations of the students. Moreover, the feedback got from a formal assessment is but grades, which do not provide any idea about what the students’ weaknesses are or which area needs reinforcement (Harris and McCann, 1994). An assessment is meant to inform teachers about their teaching practices and their students’ learning progress. Likewise, an assessment is the mirror through which students can see their abilities and achievements. In this line of thought, Blanche claims that,

Students need to know what their abilities are, how much progress they are making and what they can (or cannot yet) do with the skills they have acquired. Without such knowledge, it would not be easy for them to learn efficiently." It is easy to lose sight of the fact that tests must serve student needs as well as the teacher needs, and that an intimidating, artificial construct designed only to provide an acceptable “bell-curve” will not help the testee, and can even retard progress by demotivating him/her

(1988, as cited in Taeduck & Finch, 1998, p. 4)

However, using informal assessment (also called continuous and authentic assessment) allows the teachers to construct an idea about their real abilities and their needs in different area because they are doing this over a sufficient amount of time using several types of tasks. It also enables the teachers of profiting from the time of instruction to be an assessing time as well, thus the practicality principle is guaranteed. Harris and McCann highlight this point,

Informal assessment is particularly important for speaking as many teachers have practical difficulties in organising oral tests (lack of a place to do them or lack of time) In addition, informal assessment of speaking can have an important effect on learners. If they see that speaking English and participating in class is rewarded they will be more motivated to participate actively and try to use English in the class (1994, p. 9)

Harris and McCann (1994) suggest that teachers, using informal assessment, can evaluate their students' abilities by giving grades based on a pre-prepared rating scale. As such, they propose an example of a practical way of assessment based on assessing students' performance at a given time. They state that they have to tell students that they will be assessed, for example, during the next few sessions, then they assess a group of five or six students per session without telling them which group is being assessed.

Heaton (1997, as cited in Restrepo et al, 2003) believes that, in educational contexts, speaking performance is a specific language area that should be better assessed through informal assessment. The informal assessment helps teachers to get a clear picture of their students' ability through authentic performances (in their ordinary classes, i.e. far from the threatening tests conditions), and over a longer period of time. It also enables them of assessing some matters that cannot be assessed using formal tests such as efforts, attitudes, and perseverance (Heaton, 1997, as cited in Restrepo et al, 2003). Authentic assessment is the best ongoing process that enhances the students' motivation to learn, and to adjust both teaching and learning systematically. O'Malley and Valdez Pierce define authentic assessment as,

A type of assessment that reflects student learning, achievement, motivation, and attitudes on classroom activities. Authentic assessment is important mainly because it is based on activities that represent classroom and real-life settings and is consistent with classroom and instructional objectives.

(1996, as cited in Restrepo et al, 2003, p. 65)

Some advantageous characteristics underlie the authentic assessment and distinguish it from the traditional (formal) assessment. Brown (1998, as cited in Restrepo et al, 2003, p. 65) summarizes these characteristics as follows:

- Authentic assessment encourages students to do different things such as to perform, to create, to imitate, etc.
- Authentic assessment assesses students within a real or simulated context.
- Authentic assessment assesses what students are learning.
- Authentic assessment assesses using a variety of meaningful tasks
- Authentic assessment indicates both the strengths and the weaknesses of learners
- Authentic assessment gives learners feedback at different learning phases.

The feature of giving feedback is very important also for teachers because it shows them which area needs reinforcement, which sub-skills learners lack, and which teaching practice or even which behaviour of their parts needs to be reconsidered. As regards students, authentic assessment gets them involve in the assessment procedure, thus being responsible for their learning (Restrepo et al, 2003).

Authentic assessment can be performed using many formats. Performance-based assessment, self and peer-assessment, and Portfolio assessments are some examples.

#### **4-2-1 Performance-based Assessment**

A performance-based assessment is a means for measuring “students’ cognitive thinking and reasoning skills and their ability to apply knowledge to solve realistic, meaningful problems” (Hammond & Adamson, 2010, p. 7). A performance assessment is based on how can the learner proves his or her competence against given standards (O'Malley & Pierce, 1992). Brown points out that,

Performance-based assessment, involves **productive, observable** skills such as speaking and writing, of content valid tasks. Such performance, usually, but not always, brings with it an air of authenticity real-world tasks that students have had time to develop (2004, p. 255)

Assessors should determine the purpose behind the assessment, identify the convenient criteria, get students aware, and use a reliable evaluation form (Brown, 2004).

Oral performance can take several shapes. It can be an oral interview, storytelling, simulations, role-play, etc. The teacher observes these performances and evaluates them using the relevant an observation checklist (O'Malley & Pierce, 1992). However, it is so problematic to assess a class with a huge number of learners at a time. A solution for that is to divide the learners in small groups of five or six members (it depends on the time available, the type of the task, and the experience of the teacher), and focus only on one group at a time. It is preferable to record their performance in order to assess them at ease and in the relevant time. This way provides the teacher a permanent record for the learner's portfolio. The remaining learners can be occupied by performing self and peer assessments; this is also useful for their portfolios (Ioannou-Georgiou and Pavlou, 2003).

#### **4-2-2 Self and Peer-Assessment**

When students assess themselves, they become responsible for their learning and autonomous. This responsibility and autonomy push them to reinforce their strengths, firm up their weaknesses, and plan for their own learning to reach their goals. Self and peer-assessment enable the teachers alike of assessing their teaching practices, thus to indicate what needs change or adaptation for the sake of improving teaching and learning as well (Restrepo et al, 2003).

Harris and McCann (1994) advise that when it comes to peer-assessment to assess a speaking performance it is better to use a pair or a group performance rather than individual performance, such as role play or simulation. It is better to rely on pre-determined criteria for assessment in this case to get an instant feedback. This can be set for by assigning a group of students to do the task (a role-play for example) and another group to assess them, and then they exchange the roles. Using technical recording devices (such as tape recorder, camera, etc.) is recommended in such tasks because they have the advantage of reviewing the performance at any moment. Another advantage of using these devices is to use them as a proof that serves to measure the progress of the students and to compare their performances at different phases. Regarding scoring, teachers are advised not to ask the students to give marks to themselves; rather, it would be sufficient to ask them just to tell about their weaknesses and their strengths.

Self and peer-assessment get the students involved in the assessment process, as when they evaluate themselves they are put in a position of responsibility (Harris and McCann, 1994). Gipps (1994) points out that self-assessment enhance intrinsic motivation and intrinsic



motivation enhance self-assessment. Moreover, self- assessment suits more those learners lacking self-confidence in that they, usually, do not ask for help in their learning (Hattie & Timperley, 2007, as cited in Gonzales, 2012).

Regarding the issue of reliability in self and peer-assessment, Stefani (1994, as cited in Gonzales, 2012) states that a high correlation exists between teachers' assessment and self and peer-assessment. Besides, students conveyed that they had acquired a sense of autonomy more than that when they were assessed only by the teacher. Another study, conducted by Hughes and Large (1993), confirmed that the results obtained from the teacher assessment and those from peer-assessment were comparable (as cited in Gonzales, 2012).

### **4-2-3 Portfolio Assessment**

Portfolio assessment is a set of students' classroom work gathered over time. The purpose of this work collection is to record and evaluate the students' progress during this period (Restrepo et al, 2003).The essential components of portfolio assessment include "Collection, Reflection, and Selection" (Hamp-Lyons and Condon 2000, as cited in Douglas, 2000, p. 75). Students, when reflecting on their portfolios, become more autonomous in their learning process; they become able to construct their own progress, "A portfolio, the purpose of which is to foster learning over time, is based upon a constructivist model" (Wilkerson, 2004, as cited in Özdemir-Çagtay, 2012, p.4).

O'Malley & Pierce, (1992) state that there is no prescribed method to design portfolios, nevertheless, teachers in each educational institution can proceed to plan for their own portfolio applying the following steps:

- 1- Determining the portfolio' s purpose depending on the learning goals;
- 2- Classifying the relevant tasks and tools;
- 3- Indicating the adequate criteria;
- 4- Engaging students and their parents in this process

By accomplishing these steps, teachers will have realized the hardest part of the design. The next step will be to put them into practice. Students also have to share in this assessment; they should perform a systematic written reflection on their works, as well as assess their peers' works by giving oral feedback. These help teachers in the scoring procedure (Özdemir-Çagtay, 2012).

Oral portfolios are likely to get the students aware about their oral ability, as well as constructing their own development within a collective classroom work. Furthermore, students can overcome their anxiety when they observe their progress. Many empirical studies confirm this idea. Wang and Chang (2010) work highlights the effectiveness of self-monitoring and oral portfolio in decreasing the students' anxiety (as cited in Özdemir-Çagtay, 2012). Baron and Boschee (1995) suggest that in order to benefit from oral portfolios in reducing students' anxiety level, teachers should enhance the students' achievement' spirit (as cited in Özdemir-Çagtay, 2012).

Teachers can use this technique using different devices based essentially on technology since the speaking skill cannot be recorded only by the human memory or technical devices. Özdemir-Çagtay (2012) points out that technology-based portfolio helps overcoming those problems stemmed from the traditional assessments such as subjectivity, and affective difficulties (anxiety and motivation). Thanks to technology, oral portfolios are now easier to be used as an effective method for teachers to gather more information about the students' progress and for students to listen and see themselves when performing, and thus see their gaps and regulate them. Many tools are available to use such a portfolio. Teachers and students alike can use computers, audio and video-recorders, MP3 players, Mp4 players, and mobile phones to record oral performances.

The oral portfolio requires several types of technical aids that are not always available in our university. However, if comparing with the limitations of traditional assessments, one would rather be in favour of it. Özdemir-Çagtay (2012) states that "the majority of the pupils surveyed... expressed that the positive sides of an oral portfolio overwhelm its flaws" (p. 104). It is easy to get round the issue of lack of material, but it would be very difficult to regain our students' capacities and ambitions if they were lost.

### **4-3 Rating Criteria**

To ensure valid and reliable assessments, teachers have to put in mind the following considerations when choosing the rating criteria:

- They should choose the convenient rating criteria relying on the definition they have given to the construct 'speaking'.
- They should all (i.e. the teachers belonging to the same institution) agree on the same range of criteria, otherwise a problem of reliability would arouse. Harris

and McCann point out that using different criterion is a source of unreliability, and therefore “it is important firstly to agree on criteria for assessing students with our colleagues. The following step is to make sure that we all agree on exactly what the assessment scales mean (1994, p. 10).

- To ensure a more valid assessment, the learners must be aware of these criteria in advance to get a clear picture of what will be judged and how to arrange for it (Weir, 2005).

Having completed these conditions, teachers can assess their students at any time they want as they can give them to their students to use them in a self-assessment (Harris and McCann, 1994).

Nowadays, many common models of rating scales are available. Teachers can select, among them the closest model to their context and use it. The CEFR, for instance, though being a holistic rating, is the scale the most accredited by many universities across the European Universities. The following table is extracted from a CEFR grid for a speaking self-assessment, which includes six levels of ability (A1, A2, B1, B2, C1, and C2):

		A1	C2
S P E A K I N G	Spoken Interaction	<i>I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say.</i>	<i>I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely.</i>
	Spoken Production	<i>I can ask and answer simple questions in areas of immediate need or on very familiar topic</i>	<i>If I do have a problem I can back track and restructure around the difficulty so smoothly that other people are hardly aware of it</i>
		<i>I can use simple phrases and sentences to describe where I live and people I know.</i>	<i>I can present a clear, smoothly flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points</i>

**Table 4.1. Common Reference Levels: self-assessment grid**

**(Council of Europe, 2003, p. 26)**

The CEFR provides also a rating scale that involves the following qualitative criteria of spoken language use: Range, Accuracy, Fluency, Interaction, and Coherence. Range refers to the vocabulary repertoire that the testee uses when speaking. Accuracy means to use correct grammar and pronunciation. Fluency reflects the ability of speaking in a normal speed without much hesitation. Interaction refers to the ability of communicating with others by initiating topics, taking turns whenever relevant, sharing opinions, and maintaining a conversation. Coherence implies the logical construction of utterances using the adequate markers (Varela & Palacios, 2013).

Varela and Palacios (2013, p. 62) introduce some example of rating criteria that are used in different test types. The Trinity London Exams use as criteria: *coverage of communicative skills appropriate for the grade, language functions, grammatical, lexical and phonological items, accuracy and appropriateness of language use, and fluency and promptness of response*. The Pearson Tests of English (formerly London Tests of English) assess the following criteria: *fluency, interaction, range, accuracy, phonological control, sustained monologue, thematic development, sociolinguistic appropriateness, and turn taking*.

#### **4-4 Ensuring a Positive Washback Assessment**

The effect of assessments on learning and teaching is a crucial factor that teachers should pay attention to. The focus of this study is to administer a kind of assessment that has a positive washback on learners' speaking ability. Hence, "If the aim is to promote oral skills in the target language, then it seems reasonable that this ability should be at the heart of the testing, and that any test-driven learning that takes place should be directed towards such a goal" (Taeduck & Finch, 1998, p. 2). Teachers have to deliver assessments that really assess their students' speaking skill, not some intermittent utterances, or some sentences that are learned by heart. Similarly, if teachers focus on the linguistic rather than the communicative competence, students will focus on the linguistic competence to pass the test. For instance, "An over-reliance on grammar tests gives students the clear message that they have been wasting their time trying to communicate in class. What matters is grammar" (Harris and Macacan, 1994, p. 2).

Additionally, the timing and the purpose of the assessment has its effect on learners. Formative assessments are more fruitful than summative ones since they provide the learners with a feedback on their achievement at the right time before being too late. Brown (2004)

comments on this point saying that informal performance assessments have a positive washback because students receive an instant feedback on their performance so that they know their gaps or their strengths. However, formal assessments provide no washback if the feedback is but a letter or a numerical score. Brown adds that some alternatives in assessment such as self-assessment can have also a positive washback since it challenges the students to adjust their learning process.

Hughes (2003) says, “If you want to encourage oral ability, then test oral ability” (p. 53). Hughes argues that it is not sufficient to test only such an ability to get a beneficial washback; rather, it should be given a sufficient weight compared with the other abilities. For example, if the test purpose is to enhance students’ interactive skill, the assessor must assign the highest weight to this skill in order to get the students give more interest to this skill. Hughes (2003) provides a set of instructions that help ensure a beneficial washback:

- Test the target skill directly using as authentic tasks as possible in order to give the students the incentive to practice that skill in the way we want them to do.
- Use criterion-based assessments, which are more beneficial than norm-referenced ones. Students will learn to improve their level rather than achieving only the minimum ceiling set for success.
- Focus on the objectives of the test.
- Make the test instructions as clear as possible.
- Pay attention to the test practicality. Teachers afford not to waste efforts and time in non-useful tests.

In this vein, Muñoz et al. (2003) reports that a group of researchers at a language centre in Colombia, South America conducted a study examining the washback of classroom-based assessment. The group developed an Oral Assessment System (OAS) that consisted of:

- A set of criteria describing different aspects of oral language, namely communicative effectiveness, pronunciation, grammar, vocabulary, and task completion
- Speaking standards for each course
- A list of assessment tasks for each course
- A scoring scale
- A card reporting feedback on midterm and final assessment for the students
- Oral Assessment Guideline that explains in detail the system.

The results brought about by this system (OAS) showed a positive washback deriving from:

- The fact that students were well informed about all the OAS details, including the assessment objectives and the rating criteria, which pushed them to concentrate their learning on specific goals.
- The correlation between instructional and assessment goals encourage the teachers and the students alike to work diligently to achieve these goals.
- The use of self-assessment that helps the students to control their own progress using their own internal criteria and become more goal-oriented, thus more autonomous.

At last, one can say that this study is a source of appreciated evidence on the positive washback of an ongoing process of assessment of students, teachers, and curricula. Hence, sound decisions can be made to change the educational system (Muñoz et al., 2003).

#### **4-5 Speaking Assessment Tasks**

For the sake of ensuring a test validity, the assessment tasks should be similar to those performed in the classroom. Thornbury (2005, p. 90) provides some criteria that must underlie the classroom activities in order to boost the learner autonomy and motivation. These criteria are:

**Productivity:** Productive tasks push the learners to make efforts looking for words and structures.

**Purposefulness:** The more the tasks the learners determine a purpose for their talk the more their productivity increases

**Interactivity:** Tasks that do not involve any kind of interaction are not said to be real-life tasks and are not useful

**Challenge:** A challenging task enhances the students' enthusiasm and push them to speak and show their opinions; however, the teacher has to be careful to the degree of this enthusiasm, otherwise the students may resort to their L1.

**Safety:** Teachers have not to embarrass learners if they make errors when speaking. If they keep interrupting them to correct their mistakes, they obstruct their fluency and may be their motivation to speak.

**Authenticity:** Authentic tasks make students feel at ease, and spontaneous because they find themselves communicating as if they were outside the classroom. The selection of tasks must cope with the students' needs and interests and with their social context.

The results derived from the present research work reveal that a majority of students like better collective, interactive tasks. On this ground, teachers have many possibilities to choose a variety of tasks that meet the different students' demands and the conditions listed above as well. Thornbury (2005, p. 94) suggests a range of tasks that might be useful for this sake. Some of these tasks are the following:

**a- Stories, jokes, and anecdotes,** which are very common in our daily life. This can be monologues or dialogues. Here are some ideas to get this type more interesting:

- **Guess the lie:** A learner tell each other three anecdotes; two of which are true and one is untrue, but plausible. The remaining learners have to guess which one is the untrue and give the reason for their answer.
- **Insert the word:** Learners are given cards with an unfamiliar word or expression and are required, one-by-one, to tell a short story that includes this word; the listeners at the end of each story have to estimate what that word was.
- **Chain story:** All learners share in telling an extemporaneous story in a chain way; i.e. a learner starts telling the story and stops at a given sign from the teacher, then another learner continue telling it on the basis of what his or her friend has said and so on until the end.

**b- Drama, role-play, and simulation:** Drama tasks enable learners of using a real-life English where a wide range of vocabulary that is not usually found in the classroom can be practiced. They "provide a useful springboard for real-life language use" (Thornbury, 2005, p. 96). A role-play involves acting as another person such as pretending to be a doctor treating his patients, however, in a simulation; learners imagine themselves in different situation. For instance, they may act as if they are arranging a party, or calling their friends to get an after-dinner walk, etc. the term *drama* stands for the three types and for other types (Thornbury, 2005).

Brown (2004) states that a role-play task is advantageous in that it elicits the linguistic creativity of the learners, and lowers their anxiety because when performing, they take on the

persona of someone else. Using such a task enables the examiner to uncover the speaker's oral abilities for it is a reproduction of a real world.

**c- Discussions and debates:** the best classroom discussions and debates are those underlined by spontaneity. However, this is not always possible, so the teacher has to plan for this type of activities. Brown (2004) points out that this type of tasks serves to elicit several abilities such as attention getting, interrupting, clarifying, questioning, paraphrasing, comprehension signal like nodding, politeness and other sociolinguistic factors. He claims that this task fits more assessments that are informal. The following tasks are some examples of discussions and debates:

- **Discussion cards:** the teacher prepares in advance a number of topics, depending on the number of groups, written on cards. Each group will discuss its topic, then prepare a summary of what has been said to discuss it with the whole class. The choice of the topic can be given to the group.
- **Warm-up discussions:** Before introducing any topic, a warm-up step is necessary to draw the attention of the learners. This warming can be a set of questions, a picture, a brief introduction, etc.
- **Balloon debate:** A common debate where the discussion is about a hot air-balloon, which is in danger because of its excess load of passengers. To get out of this impasse, at least one of them must be thrown. Each passenger, being a famous, living, or historical person, attempts to offer the appropriate justifications to convince the others of the need to keep him or her alive.
- **Pyramid (or consensus) debate:** the norm for this discussion is to get all the class agree on a given issue. The issue can be for example, drafting an internal law for the classroom consisting of ten articles, or ideas like "*the eight necessary things to bring with us to a desert island*", or "*the ten most important persona in the world*". It begins with a pair of students trying to convince another pair. Then, each pair goes to convince again another pair, and so on, until the entire class arrives at a consensus.
- **Panel discussion:** this simulates the format of a television debate where few learners representing different directions discuss and defend their viewpoints in a presence of a chairperson; the remaining learners represent the audience and can ask questions. Panel discussion is very successful if the learner represents really his or her point of view.



**d- Conversation and chat:** Conversation means the casual talk that occurs in the classroom between the teacher and the students. It is the most common speaking aspect, which “is not the result of language learning, but it is the site where learning occurs” (Thornbury, 2005, p. 105). Most learners need to develop their conversational interaction, which, is a spontaneous and an unstructured task. Hence, a conversation task has to be delivered to the learners, including conversational expressions (such as how to open or close a conversation, how to interrupt, how to change the topic, etc.) and communication strategies (such as paraphrasing and avoidance strategy). An example that encourages this type of tasks is to bring a guest to the classroom and start a free talk with him or her.

**e- Oral Interviews:** Brown (2004) states that oral interviews are the first task thought about when it comes to a speaking assessment. However, these interviews are usually performed in a way that do not include a natural interaction, i.e. the two partners have not a similar status that creates similar opportunities in turn taking. The examiner, being at a high status, only asks questions, and sometimes challenges the learner and this latter has to answer. Weir sheds the light on this point saying:

Clearly, if we wish to test spoken interaction, a valid test must include reciprocity conditions. This contrasts with the traditional interview format in which the interviewer asks the questions and the interviewee answers. So if we are interested in the candidate’s capacity to take part in spoken interaction, there should be reciprocal exchanges where both interlocutor and candidate have to adjust vocabulary and message and take each other’s contributions into account. This means that the candidate must be involved in the interaction to a greater extent than that of merely answering questions (2005, p. 71)

Weir asserts that if we want to get a valid test, then we have to respect the ‘*reciprocity condition*’, that is to say that both partners have the same speaking rights during the interview. This, contrasts with what is being occurred in traditional interview format where “the examiner dominates the conversation at least in terms of agenda management: initiation of discussion, continuance and completion” (Weir, 2005, p. 71)

Therefore, if we have to perform an oral interview, it is necessary to assess two or more learners at once. This is sufficient to ensure a '*reciprocity condition*', thus ensuring interaction and a low level of students' anxiety.

#### **4-6 Lessening the Difficulties Encountered by the Students**

The present research work has uncovered the difficulties the students were suffering from either in speaking learning or in speaking assessment. In this section, the researcher attempts to provide some remedies. These problems are lack of vocabulary, the personality of the teacher, and anxiety.

##### **4-6-1 Students' Lack of Vocabulary**

To speak a language, it is necessary to learn its vocabulary. Learning vocabulary is more important than learning grammar because "if you spend most of your time studying grammar, your English will not improve very much. You will see most improvement if you learn more words and expressions. You can say very little with grammar, but you can say almost anything with words" (Thornbury, 2002, p. 19).

Many learners complain of '*lack of vocabulary*' problem and teachers still deny the importance of a communicative teaching method in extending the vocabulary store of the learners. In this vein, proponents of 'vocabulary acquisition' suggest two key developments to overpower the dominance of the grammar approaches: lexical syllabus and lexical chunks. Lexical syllabus focuses on the words with high frequency in written and spoken language. Lexical chunks are meaningful units constituted of more than one word such as *look after*, *bits and pieces*, *put on*, etc. They are useful in language use and in getting fluency (Thornbury, 2002).

English words are divided into three types: words we do not need to use them, words we need to understand them when reading or listening, and words we need to use ourselves. English has many words that are rooted from other languages such as Arabic, Latin, and French. This feature makes it easy to learn and memorize a large range of words. It is not enough to know the meaning of a word in isolation; rather we have to know its meaning within a sentence, the words usually associated with it, its pronunciation, the preposition that goes with it, and its grammatical characteristics (McCarthy & O'Dell, 1994).

However, the common learners' problem with vocabularies 'to forget' the new items they learned. In fact, learning is meaningless without *memorising* what has been learned. To do that, learners have to know how memorisation is occurring, and how the human memory works. Scientists in this field state that there are three kinds of memories: the short-term store, the working memory, and the long-term memory. Hence, different exercises should be done to save our words in the *long-term memory*. This process of saving is the business of the *working memory*. The information is first located in this memory, analysed, and classified to be retrieved later on. However, storing new items in the *long-term memory* does not mean that they will be there forever; they would be forgotten if they were not retrieved many times. This retrieval can be done through many practices such as repetition, spacing, use, mnemonics, and many other tricks (Thornbury, 2002).

To conclude with, practical pieces of advice should be provided to the learners to help them build an extended range of vocabulary. McCarthy and O'Dell (1994, p. 2), suggest the following practices:

- When reading or listening to English do not focus on every single word or utterance, but try to get the overall meaning then you can understand the meaning of some words.
- You have to plan to your vocabulary learning: how many words by week and where and how to learn them.
- Learn each new word with its associated words, (e.g. a cat, kitten, mew, paw), and compare your work with that of your fellows to add other words to yours.
- Pictures and drawings help a lot in remembering the learned words.
- Organize words, in a notebook, depending on their meanings. Each learner has his or her own method to do this. For example, the word '*furniture*' includes many types; it is better to categorize each type together:
  - *Living room*: an armchair, a bookcase, a carpet, a sofa, a television.
  - *Bathroom*: a basin, a bath towel, soap, hairbrush, shampoo, etc.
  - *Bedroom*: a pillow, a duvet, a wardrobe, a drawer, a stool, a mirror, etc.

Smith (2012) recommends a way to learn vocabulary by knowing word parts, which are *prefixes*, which are added at the beginning of a word; *suffixes*, which are added at the end of a word, and *root words*, which are the essential part of a word. However, we have to be careful;

this rule is not applicable for all the words. Smith, also, advises learners to learn and be familiar with the *challenging words* and *confusing words*. Challenging words are the difficult words that appear usually in textbooks, newspapers, periodicals. Confusing words are words such as affect/ effect, whether/weather, council/counsel.

#### 4-6-2 The Personality of the Teacher

It is reasonable and widely known that the personality of the teacher plays a crucial role in students' academic achievements. It can restore the pent-up capacities of the student, as it can suppress the student's will of learning. The teacher's personality includes two kinds of qualities: personal and experiential. *Personal qualities* refer to traits such as aptitude, age, and gender. *Experiential qualities* mention the professional experiences the teacher have, such as qualification, educational background, and previous experience in teaching (Ashton, as cited in Kosgei et al, 2013).

A study made by Ferguson () mentioned that *teacher's qualification* plays a critical role in students' achievement and that "every additional dollar spent on more highly qualified teachers netted greater increases in student achievement than did less instructionally focused uses of school resources" (1991, as cited in Darling-Hammond, 1999, p. 12). Wilson et al (2001, as cited in Kosgei et al, 2013) assert that the more certified and well prepared the teacher, the more successful the relationship with his or her students. A good relationship between the students and their teacher foster their motivation and their academic achievement (Davis et al, 2006, as cited in Garcia et al, 2011).

Concerning the teacher's *knowledge of the subject matter*, many studies confirmed the positive relationship between it and the students' success (Goldhaber & Brewer, 2000, as cited in Kosgei et al, 2013). However, Darling-Hammond (1999), relying on a number of studies (Andrews, Blackmon & Mackey, 1980; Ayers & Qualls, 1979; Haney, Madaus, & Kreitzer, 1986; Quirk, Witten, & Weinberg, 1973; Summers & Wolfe, 1975) states that the knowledge of the subject matter has a slight impact on the students' academic achievement.

*The experience of the teacher* is also a factor that has an influence on the students' learning. Stringfield & Teddlie (1991, as cited in Kosgei et al, 2013) argue that the more experienced teachers use appropriate ways to convey the information to their students. Whereas Rosenholtz (1986, as cited in, Darling-Hammond, 1999) reveals that the teacher's experience stops to increase after about five years of teaching because, perhaps, "older teachers do not always continue to grow and learn and may grow tired in their jobs" (p. 10).

Rosenholtz adds that beginning teachers if they were well prepared can be very effective as well.

Other studies refer to the teacher's *verbal ability* to be a decisive factor for students' achievement. Stronge et al. (2007, as cited in Kosgei et al, 2013) believes that the verbal ability, as being evidence of teacher quality, has a positive relationship with students' achievement. Murnane (1985, as cited in Darling-Hammond, 1999) explains that the teacher having a verbal ability can transfer the information persuasively.

Apart from what have been said about the experiential qualities of a teacher, his or her personal qualities are considerable too. Moscoso and Slagado (2004, as cited in Garcia et al, 2011) examine the effect of the dark side of the personality on job performance. The results from this study show that seven personal characteristics prevent teachers from doing their job effectively. These characteristics are shyness, suspiciousness, sadness, pessimism, suffering, eccentricity, and riskiness. Darling-Hammond (1999) reports that many studies approve that a positive relationship exists between students learning and teachers' flexibility, creativity, or adaptability. Cooper and Benis (1967 as cited in Garcia et al, 2011) suggests that teacher's behaviour should be regulated to get learners learn. Zarabian et al (2015) report that the teacher's personality can include good qualities that enhance the students' motivation or bad qualities that decrease it.

As regards *the communicative skill*, little attention was paid to the impact of the teacher's personality on student's communicative ability (Hashim et al, 2014). Hashim and others conducted a study in 2014 in a Malaysian institution investigating this issue. They define the personality of the teacher as being a sum of inner qualities, which can be upraised through expression, attitude, and behaviour (stronge et al, 2004, as cited in Hashim et al, 2014). As for the attributes of a good teacher, they reported that a teacher should be skilled in their job as well as in their personal qualities. Personal qualities include traits such as being thoughtful, fair, respectful, having positive attitudes towards teaching, and having social interactions with students. "...attractive teacher personality, apart from teaching techniques and materials, is critical to EFL learning" (Hashim et al, 2014, p. 103). Schneider suggests that "respectful teachers know their students by name, value each individual's ability, and are aware of student's frame of mind" (Schneider, 2002, as cited in Hashim et al, 2014, p. 105). Hashim and others believe that building a social connection between the teacher and the students participate in creating a good learning atmosphere and decrease the students' anxiety.

Concisely, Hashim and others' study reveals a strong relationship between teacher's personality and students' communicative skill. The more positive the personality of the teacher the more the communication skill of the students increases and anxiety decreases. As such, it is better for the teacher to combine both experiential and personal qualities.

#### **4-6-3 Students' Anxiety**

Anxiety, in general, is defined as "as experience of general uneasiness, a sense of foreboding, a feeling of tension" (Hensen, 1977, as cited in Wilson, 2006, p. 25). Hence, language anxiety is the negative feelings of nervousness, discomfort, and apprehension that go with language learning. Anxiety might exercise a harmful impact on language learning, and anxious students "will have lower levels of verbal production ... and will be reluctant to express personally relevant information in a second-language conversation" (Gardner and Mac Intyre, 1999, as cited in Djebbari, 2014, p. 51).

Tallon (2008, as cited in Pršić, 2013) states that language anxiety is stemmed generally from three main sources: learner's characteristics, teachers' characteristics, and classroom's characteristics. Young (1991, as cited in Wilson, 2006) suggests six possible sources for language anxiety: "(1) personal and interpersonal anxieties, (2) learner beliefs about language learning, (3) instructor beliefs about language teaching, (4) instructor learner interactions (5) classroom procedures, and (6) language testing" (p. 78).

Instructor beliefs, such as being intimidating, correcting every single mistake, and group work is out of control, determines the instructor-learner interactions and arises the students' anxiety (Young, 1991, as cited in Wilson, 2006). Palacios also refers to the role of the teacher in increasing students' anxiety,

Absence of teacher support, unsympathetic personalities, lack of time for personal attention, favouritism, absence that the class does not provide students with the tools necessary to match up with the teacher's expectations, and the sense of being judged by the teacher or wanting to impress the teacher (1998, as cited in Pršić, 2013, p. 22)

As for practicing the four skills, speaking is the most anxiety provoking skill that many language students suffer from (Wilson, 2006). Speaking means taking the risk of making errors in front of the others, having bad marks and failing in the exam, exposing oneself to the

criticism of the teacher, or getting embarrassed (Beebe, 1993, as cited in Djebbari, 2014). Speaking in front of the classmates without being prepared is also a source of anxiety (Young, 1991, as cited in Pršić, 2013). Leary (1982) views that the learner's perception that speaking should be performed as perfect as possible can be a source of anxiety (as cited in Wilson, 2006). Horwitz et al. (1986, as cited in Pršić, 2013) relate anxiety to learners' insufficient communicative competency in the foreign language.

In order to alleviate speaking anxiety in foreign language classrooms, teachers have to pay attention to the aforementioned resources of anxiety and make efforts to help the students overcoming this disturbing psychological state. It is necessary for the teachers to choose carefully the tasks that are not threatening. Pair and group work can be a remedy for that problem, however, speaking individually is not recommended unless the students are ready to do it. Another condition is the authenticity and contextualization of the tasks. Assessment tasks should simulate the real-life activities and involve familiar and interesting contents for the learners (Ioannou-Georgiou and Pavlou, 2003).

Hembree (1988, as cited in Pršić, 2013) suggests three approaches to alleviate students' anxiety: the cognitive, the affective, and the behavioural approach. The cognitive approach tends to change the student's self-evaluation to be more positive. The affective approach attempts to remove the link that the students make between classroom and anxiety. Students are taught how to be relaxed in the classroom. Using the behavioural approach, the students acquire the behavioural skills that help them to be successful in particular oral communication settings. Young states,

(1) for anxieties stemming from learner's personality, providing supplemental instruction or a support group; (2) for anxieties stemming from classroom procedures, using more pair and group work; (3) playing language games with an emphasis on problem-solving; and (4) role-playing with preparatory activities to instil class rapport. The students felt more at ease when the instructors' manner of error correction was not harsh and when they were friendly, patient, and had a good sense of humour.

(1990, as cited in Pršić, 2013, p. 23)

Accordingly, teachers have the greatest responsibility of sharing in alleviating their students' anxiety, especially when it comes to the practicing the speaking skill either in the classroom or in the periods of examinations because this skill is the most anxiety provoking

skill as many researchers acknowledge. Moreover, a carefully studied oral assessment-training programme for the teachers of the concerned module becomes an urgent need to reform the current situation and to give the speaking skill its real value.

#### **4-7 Conclusion**

Relying on the findings of this research work, this chapter has tried to propose several remedies for the complications that impede the achievement of successful speaking assessment. It has suggested the appropriate type of assessment, the relevant rating criteria, and the convenient tasks for this intricate skill. It has also shed the light and recommended some cures for the difficulties that hamper the students' speaking ability such as the lack of vocabulary, the negative side of the personality of the teacher, and students' anxiety.



## **General Conclusion**

The assessment of speaking, as being the most complicated skill, has long been a controversial subject in the field of language assessment. The different definitions scholars provided to the construct '*speaking*' brought about different grounds to assess a learner's ability to speak. This research work tried to highlight this issue within the English section of the University of Tlemcen. The major goal of the researcher was to find out the teachers' perception of *speaking* and the way of assessing their students' speaking skill, thus diagnose the deficiencies of the actual speaking assessment practices, and describe the appropriate remedies and adjustments.

To reach this goal, the dissertation was divided into four chapters. The first chapter was devoted to review the literature concerning the speaking assessment and its principles as well as its challenges. The aim of such a theoretical research was to find out a standard or point of reference against which the actual assessment practices can be compared or assessed.

The second chapter was dedicated to describe the educational situation at the three levels of English teaching in Algeria (middle, secondary, and tertiary levels) as well as to explain how the researcher proceeded to conduct this study in terms of methodology, sampling, and collecting data.

The researcher followed a mixed approach to research, which combines qualitative and quantitative data sources. She conducted an explanatory case study that enabled her to get a closer picture about what was being occurred and what regulations were needed. The population addressed were the second year license students, that is two hundred and forty one student, in the English department of the University of Tlemcen. Since it was hard to reach all the students of the second year, the size of the students understudy was restricted to thirty students. The research instruments used to collect data were the students' questionnaire, the teachers' interview, and the test observation. These different tools were beneficial for the researcher to get insights from different angles.

Answering the questionnaire by the students enabled the researcher of knowing the students' attitudes towards the speaking skill, the speaking assessment, and towards the practices of their teachers. It allowed the researcher also to get an insight of the main difficulties they encounter when speaking either in their classes or during tests. The interview permitted the researcher to get the facts in the teachers' eyes. They explained to her the way they proceed to manage a speaking lesson, to administer a speaking assessment, and to deal with students' problems. They spoke also about the difficulties they face when teaching or

assessing their students as well as their wishes concerning the assessment of speaking. The test observation was an opportunity for the researcher to observe real test performances and to see different students' reactions and teachers' practices.

As regards the third chapter, the focus was on the analysis and the interpretation of the collected data. Triangulation (i.e. varying the sources of data) helped the researcher to get the true insight of events and thus to analyse and interpret them objectively. The researcher arrived eventually to conclude that the assessment of speaking at Tlemcen University is not subjected to the assessment standards and that many regulations are needed to improve the teachers' practices. Concisely, the main findings of this study were as follows:

- ❖ The vast majority of the students yearn to speak the English language as natives, but they lack autonomy and they do not find sufficient opportunities to practice.
- ❖ Teachers have many difficulties either in teaching or in testing speaking.
- ❖ Tests are not practical.
- ❖ Tests have a negative washback on students' motivation.
- ❖ Students have many difficulties in speaking tests among which lack of vocabulary and the personality of the teacher are the most important ones.

Accordingly, the hypotheses put by the researcher were largely confirmed excepting a small part in the third hypothesis related to the order of difficulties facing the students when speaking. The researcher supposed that anxiety and lack of vocabularies are the main problems that hinder the students' ability to speak while the results of the study showed that the lack of vocabulary is the first hindrance to students' speaking ability then comes the personality of the teacher. As for the first hypothesis, the present study firms up that the teachers do not assess speaking as an interactive skill, they assess only the students' production of speech. The evidence for this is that teachers use a set of criteria that does not represent or reflect the interactive skill of the students. The majority of the teachers focus on pronunciation, grammar, fluency, and presentation. Likely, the second hypothesis was strengthened too, since teachers were still using the same test task, which was the face-to-face one-way interview to assess their students' performances.

The fourth chapter was offered to help teachers and students alike to overcome the detected problems and difficulties that presented an obstacle to promoting the speaking assessments and improving the students' performances. Concerning the first hypothesis, the

researcher provided the teachers with pieces of advice related to the criteria of speaking assessment that should be used. As for the second hypothesis, the type of assessment suggested to get more objective rating and to motivate students to make more efforts was the authentic assessment. In other words, the students' speaking ability should be better assessed during their ordinary classes, far from the threatening and intimidating conditions of tests. This can be done through many faces of authentic assessment such as portfolios and self and peer-assessments. As such, the researcher advocated the use of authentic tasks, i.e. the tasks that involve more interaction and that enhance the students to perform in a more relaxed state. As regards the third hypothesis, the current study advised the teachers to be more effective to assist their students to improve their speaking ability and to administer assessment and testing that are more valuable. Likely, it offered some strategies for the students to extend their vocabulary range to be more fluent and to get rid of anxiety.

This research work represents a first step towards change and innovation in the field of language assessment. The English language, in this third millennium is the key to all kinds of relationships and communications between our country and the worldwide countries. Hence, it is compulsory to reconsider the traditional educational methods and practices being followed in our universities in teaching and assessing this language in general and the speaking skill in particular. The students need to speak and understand the spoken English; that is to use the language in real life situations not only within the confines of classrooms. For this reason, the current research work is an attempt to get the attention back to this vital skill, which is not given its real value and importance; it is a call for further researches that provide more details about the assessment of the learners' speaking ability.

## **BIBLIOGRAPHY**

### **Books**

- Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. London: Continuum.
- Allen, V. F. (1983). *Techniques in Teaching Vocabulary*. Oxford: Oxford University Press.
- Anderson, G., & Anderson, G. J. (1998). *Fundamentals of Educational Research*. UK: Falmer Press.
- 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* (Vol. 1). Oxford: Oxford University Press.
- Bailey, K. M. (1999). *Washback in Language Testing*. Princeton, New Jersey: Educational Testing Service.
- Beins, B. C., & McCarthy, M. A. (2012). *Research Methods and Statistics*. New York: Pearson Higher Ed.
- Bell, J. (2005). *Doing Your Research Project: A Guide for First-Time Researchers in Education, Health and Social Science*. UK: McGraw-Hill Education. Open University Press
- Blaxter, L. (2006). *How to Research*. UK: Open University Press.
- Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. London: Longman.
- Brown, J. D. (1996). *Testing in Language Programs*. New Jersey: Prentice Hall Regents, Upper Saddle River.
- Bygate, M. (1987). *Speaking*. Oxford: Oxford University Press.
- Callies, M., & Götz, S. (Eds.). (2015). *Learner Corpora in Language Testing and Assessment* (Vol. 70). Amsterdam: John Benjamins Publishing Company.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). *Communicative Competence: A Pedagogically Motivated Model with Content Specifications. Issues in Applied Linguistics*. California: Regents of the University of California.
- Chapelle, C. (2003). *English Language Learning and Technology: Lectures on Applied Linguistics in the Age of Information and Communication Technology* (Vol. 7). Amsterdam: John Benjamin Publishing Company.

- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education*. UK: Routledge.
- Council of Europe. (2002). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Darling-Hammond, L., & Adamson, F. (2010). *Beyond Basic Skills: The Role of Performance Assessment In Achieving 21st Century Standards of Learning*. Stanford: Stanford Center for Opportunity Policy in Education (SCOPE), Stanford University, School of Education.
- Davidson, F., & Lynch, B. K. (2002). *Test craft: A Teacher's Guide to Writing and Using Language Test Specifications*. USA: Yale University Press.
- Dontcheva-Navratilova, O., Povolná, R., Aijmer, K., Hüttner, J., Malá, M., Cap, P., & Miššíková, G. (2009). *Coherence and Cohesion in Spoken and Written Discourse*. Cambridge: Cambridge Scholars Publishing.
- Douglas, D. (2000). *Assessing Languages for Specific Purposes*. Cambridge: CUP.
- Douglas, D. (2014). *Understanding Language Testing*. London: Routledge.
- Duff, P. (2008). *Case Study Research in Applied Linguistics*. London: Taylor & Francis
- Elder, C., & Davies, A. (2001). *Experimenting with Uncertainty: Essays in Honour of Alan Davies* (Vol. 11). Cambridge: Cambridge University Press.
- Fraser, B. (2010). *Pragmatic Competence: The Case of Hedging*. UK: Emerald Group Publishing Limited
- Fulcher, G. (2003). *Testing Second Language Speaking*. New York: Pearson Education.
- Fulcher, G. (2010). *Practical Language Testing*. UK: Hodder Education. An Hachette UK Company.
- Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment*. London: Routledge.
- Fulcher, G., & Davidson, F. (2012). *The Routledge Handbook of Language Testing*. London: Routledge.
- Gerring, J. (2006). *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.
- Gillham, B. (2000). *Case Study Research Methods*. London: Continuum
- Gipps, C. V. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. UK: Psychology Press.
- Griffiee, D. T. (2012). *An Introduction to Second Language Research Methods: Design and Data*. USA: TESL-EJ Publications, Berkeley.
- Harlen, W. (2006). *The Role of Teachers in the Assessment of Learning*. Pamphlet produced by the Assessment Systems for the Future projects (ASF).UK: Assessment Reform group, 347-364
- Harmer, J. (1991). *The Practice of English Language Teaching*. London.

- Harris, M., & McCann, P. (1994). *Assessment*. Macmillan Heinemann English Language Teaching. UK: Macmillan Publishers Ltd
- Heck, R. H. (2004). *Studying Educational and Social Policy: Theoretical Concepts and Research Methods*. London: Routledge.
- Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. USA: Academic press.
- Hinkel, E. (Ed.). (2011). *Handbook of Research in Second Language Teaching and Learning* (Vol. 2). London: Routledge.
- Hughes, A. (2003). *Testing for Language Teachers* (2<sup>nd</sup> Ed.). Cambridge: CUP.
- Ioannou-Georgiou, S., & Pavlou, P. (2003). *Assessing Young Learners*. Oxford: Oxford University Press.
- Jackson, S. (2015). *Research Methods and Statistics: A Critical Thinking Approach*. Wadsworth (USA): Cengage Learning.
- Kelly, G. (2000). *How to Teach Pronunciation (With Cd)*. New York: Pearson Education Limited.
- Kincheloe, J. L. (2003). *Teachers as Researchers. Qualitative Inquiry as a Path to Empowerment, 2*. London: Routledge Falmer.
- Kopriva, R. (2011). *Improving testing for English language learners*. London: Routledge.
- Kothari, C. R. (2004). *Research Methodology: Methods and Techniques*. New Delhi: New Age International.
- Kreidler, C. W. (1997). *Describing Spoken English: An Introduction*. London: Routledge.
- Kunnan, A. J. (1998). *Validation in Language Assessment: Selected Papers from the 17th Language Testing Research Colloquium*. London: Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis*. USA: Sage Publications, Inc.
- Lodico, M. G., Spaulding, D. T., & Voegtler, K. H. (2006). *Methods in Educational Research: From Theory to Practice*. San Francisco: Jossey-Bass a Wiley Imprint.
- Luoma, S. (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.
- Madsen, H. S. (1983). *Techniques in Testing*. Oxford: Oxford University Press.
- Marczyk, G., DeMatteo, D., & Festinger, D. (2005). *Essentials of Research Design and Methodology*. USA: John Wiley & Sons Inc.
- McCarthy, M., & O'Dell, F. (1994). *English Vocabulary in Use: upper-intermediate & advanced*. UK: Cambridge University Press (CUP).
- McMillan, J. H. (1996). *Educational Research: Fundamentals for the Consumer*. New York: HarperCollins College Publishers.

- Nation, I. S. P., & Newton, J. (2008). *Teaching ESL/EFL Listening and Speaking*. London: Routledge.
- Nunan, D. (1992). *Research Methods in Language Learning*. Cambridge: Cambridge University Press.
- O'Loughlin, K. J. (2001). *The Equivalence of Direct and Semi-Direct Speaking Tests*. Cambridge: Cambridge University Press.
- Pawlak, M., & Waniek-Klimczak, E. (Eds.). (2015). *Issues in Teaching, Learning and Testing Speaking in a Second Language*. New York: Springer.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: CUP.
- Redman, S. (1997). *English Vocabulary in Use: Pre-intermediate and Intermediate*. Cambridge: CUP
- Sasson, D. (2013). *Speaking and Writing for English Language Learners: Collaborative Teaching for Greater Success with K-6*. USA: R&L Education.
- Scott, D., & Usher, R. (2011). *Researching Education: Data, Methods and Theory in Educational Enquiry*. London: Continuum International Publishing Group.
- Smith, R. K. (2010). *Building Vocabulary for College*. USA: Cengage Learning.
- Susan, B. (1994). *Teaching English*. London: Routledge.
- Thornbury, S. (2002). *How to Teach Vocabulary*. New York: Pearson Education Limited.
- Thornbury, S., & Harmer, J. (2005). *How to Teach Speaking* (Vol. 156). UK: Harlow: Longman.
- Torrance, H., & Pryor, J. (1998). *Investigating Formative Assessment: Teaching, Learning and Assessment in the Classroom*. UK: McGraw-Hill Education.
- Tuttle, H. G., & Tuttle, A. (2013). *Improving Foreign Language Speaking Through Formative Assessment*. New York: Routledge.
- Walford, G. (Ed.). (2001). *Doing Qualitative Educational Research*. London: Continuum.
- Weir, C. J. (2005). *Language Testing and Validation*. UK: Macmillan.
- Wyatt-Smith, C., & Cumming, J. (2009). *Educational Assessment in the 21st Century*. London: Springer.
- Yin, R. K. (2003). *Case study research: Design and Methods*. USA: Sage publications.
- Zieky, M., Perie, M., & Livingston, S. (2006). *A Primer on Setting Cut Scores on Tests of Educational Achievement*. Princeton, NJ: Educational Testing Service.
- Harlen, W. (2006). *The role of teachers in the assessment of learning. Pamphlet produced by the Assessment Systems for the Future projects (ASF) Assessment Reform group*. UK, 347-364

**Articles**

- Al-Amri, M. N. (2010). Direct Spoken English Testing Is still a Real Challenge to be Worth Bothering About. *English Language Teaching*, 3(1), 113.
- Alduais, A. M. S. (2012). An Account of Approaches to Language Testing. *International Journal of Academic Research in Progressive Education and Development (IJ-ARPEd)*, 1(4), 203-208.
- Alonso, R. A. (2014). *Teaching Speaking: an Exploratory Study in Two Academic Contexts*. *Porta Linguarum*, 22, junio 2014, 145-160. Spain: University of Vigo.
- Bagarić, V., & Djigunović, J. M. (2007). Defining Communicative Competence. *Metodika*, 8(1), 94-103.
- Bashir, M., Azeem, M., & Dogar, A. H. (2011). Factor Effecting Students' English Speaking Skills. *British Journal of Arts and Social Sciences*, 2(1), 34-50.
- Bougandoura, F. (2012). Enhancing the Algerian EFL Learners' Speaking Skills: A Case Study. *Tlemcen University 14* مجلة الأثير، العدد 14
- Cambridge, E. S. O. L. (2011). Using the CEFR: Principles of Good Practice. *Cambridge ESOL*.
- Chelli, S. (2010). The Competency-based Approach in Algeria: A Necessity in the Era of Globalization. Mohamed Kheider University, Biskra.
- Darling-Hammond, L. (1999). Teacher Quality and Student Achievement. A Review of State Policy Evidence. *Document R-99-1*. Stanford University
- Darwesh, A. A. (2010). *Cloze Tests: An Integrative Approach*. Irak, Al-Mustansiriyah University. مجلة كلية التربية، العدد 64
- Fraser, B. (2010). Pragmatic competence: The Case of Hedging. *New Approaches to Hedging*.
- Garcia, P., Kupczynski, & Holland, G. (2011). The Impact of Teacher Personality Styles on Academic Excellence of Secondary Students. *National Forum of Teacher Education Journal*, 21(3), 1-8.
- González, A. B. (1996). Testing English as a Foreign Language: An Overview and Some Methodological Con-Siderations. *Revista Española De Lingüística Aplicada*, (11), 17-50.
- Hamad, M. M. (2013). Factors negatively affect speaking skills at Saudi colleges for girls in the south. *English Language Teaching*, 6(12), 87.
- Hashim, N. M. H. N., Alam, S. S., & Yusoff, N. M. (2014). Relationship between Teacher's Personality, Monitoring, Learning Environment, and Students' EFL Performance. *GEMA: Online Journal of Language Studies*, 14(1), 101-116.
- Hilliard, A. (2014). Spoken Grammar and Its Role in the English Language Classroom. *English Teaching Forum* N° 4.



- Idri, N. (2005). The LMD System Experience as a Struggle between the Educational Development and Reform: An Analytical Study of the Endeavour of the Academic Year 2004/2005 in Bejaia University with Suggested Solutions.
- Jeong, H., Hashizume, H., Sugiura, M., Sassa, Y., Yokoyama, S., Shiozaki, S., & Kawashima, R. (2011). Testing Second Language Oral Proficiency in Direct and Semidirect Settings: A Social-Cognitive Neuroscience Perspective. *Language Learning*, 61(3), *a Journal of Research in Language Studies*, September 2011, pp. 675–699.
- Khamkhien, A. (2010). Teaching English speaking and English speaking tests in the Thai context: A reflection from Thai perspective. *English Language Teaching*, 3(1), 184.
- Kim, H. J. (2006). *Issues of Rating Scales in Speaking Performance Assessment*. Teachers College Columbia University, 6(2).
- Knight, B. (1992). Assessing Speaking Skills: A Workshop for Teacher Development. *ELT Journal*, 46(3), 294-302. Sinwongsawat, K. (2012). Rethinking Assessment of Thai EFL Learners' Speaking Skills. *Language Testing in Asia*, 2(4), 1-11.
- KOŞAR, G. & BEDİR, H. (2014). Strategies-Based Instruction: A Means of Improving Adult EFL Learners 'Speaking Skills. *International Journal of Language Academy*, 2(3), 12-26.
- Kosgei, A., Mise, J. K., Odera, O., & Ayugi, M. E. (2013). Influence of Teacher Characteristics on Students' Academic Achievement among Secondary Schools. *Journal of Education and Practice*, 4(3), 76-82.
- Kunnan, A. J. (2000). *Fairness and validation in language assessment: selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida* (Vol. 9). Cambridge: Cambridge University Press.
- Li, W. (2011). *Validity Considerations in Designing an Oral Test*. *Journal of Language Teaching and Research*, 2(1), 267-269.
- Muñoz, A. P., Aristizábal, L. D., Orozco, F. C., Gaviria, S., Orozco, L. A. L., & Urán, M. P. (2003). Assessing Spoken Language in EFL: Beliefs and Practices. *Revista Universidad Eafit*, 39(129), 63-74.
- O'Sullivan, B., Weir, C. J., & Saville, N. (2002). *Using Observation Checklists to Validate Speaking-Test Tasks*. University of Cambridge: *Language Testing*, 19(1), 33-56.
- Oradee, T. (2012). Developing Speaking Skills Using Three Communicative Activities (Discussion, Problem Solving, and Role-Playing). *International Journal of Social Science and Humanity*, 2(6), 533.
- Pierce, L. V., & O'Malley, J. M. (1992). Performance and Portfolio Assessment for Language Minority Students. Washington, DC: National Clearinghouse for Bilingual Education. *NCBE Program Information Guide Series* (Vol. 9).
- Puhl, C. A. (1997). 'Develop, Not Judge: Continuous Assessment in The ESL Classroom'. *English Teaching Forum* (Vol. 35, No. 2, pp. 2-9).

- Rana, R. A., & Mahmood, N. (2010). The Relationship between Test Anxiety and Academic Achievement. *Bulletin of Education and Research*, 32(2), 63-74.
- Restrepo, A. P. M, Aristizábal, L. D., Orozco, F. C., Gaviria, S., Orozco, L. A. L., & Urán, M. P. (2003). Assessing Spoken Language in EFL: Beliefs and Practices. Colombia: *Revista Universidad Eafit*, 39(129), 63-74.
- Riasati, M. J. (2011). Language Learning Anxiety from EFL Learners' Perspective. *Middle-East Journal of Scientific Research*, 7(6), 907-914.
- Roca-Varela, M. L., & Palacios, I. M. (2013). How Are Spoken Skills Assessed in Proficiency Tests of General English As A Foreign Language? A Preliminary Survey. *International Journal of English Studies*, 13(2), 53-68.
- Santos, M. G. (1997). Portfolio Assessment and the Role of Learner Reflection. *English Teaching Forum* (Vol. 35, No. 2, p. n2).
- Taeduck, H., & Finch, A. E. (1998). Oral Testing and Self-Assessment-The way forward? *Andong National University*.
- Tsou, W. (2005). Improving Speaking Skills through Instruction in Oral Classroom Participation. *Foreign Language Annals*, 38(1), 46-55.
- Yazan, B. (2015). Three Approaches to Case Study Methods in Education: Yin, Merriam, and Stake. *The Qualitative Report*, 20(2), 134-152.
- Zarabian, F., Farajollahi, M., Pour, Z. Y., & Seresht, A. A. S. (2015). The Relationship between Teachers' Personality Types and Female High School Third Graders' Achievement Motivation in Mashhad. *International Research Journal of Applied and Basic Sciences*, 9 (6), 815-823.

## **Theses**

- Abderrezzag, S. (2010). *The effects of Anxiety on Students' Achievement the Case of Third Year LMD students*. Department of English University of Constantine. Master dissertation.
- Aimeur, R. (2011). *Project-Based Learning in the Algerian Secondary School Syllabuses and Textbooks* (Doctoral thesis), University of Tizi-Ouzou.
- Benmostefa, N. (2014). *Reflections upon the Baccalaureate EFL Tests as a Source of and a Means for Innovation and Change in ELT in Algeria* (Doctoral thesis, University of Tlemcen).
- Berrabah, B. (2006). *The Use of Simplified Text Versions with Close Reference to Vocabulary Teaching to EFL Learning* (Magister dissertation). Tlemcen: Abu Bekr Belkaid University.
- Berrabah, B. (2014). *Developing Vocabulary through The Receptive Skills: Case of 1st Year EFL Students at The University of Tlemcen* (Doctoral thesis), Tlemcen: Abu Bekr Belkaid University.

- Bouyakoub, N. (2012). *Exploring Writing in an Academic Context: The Case of First-Year University Students* (Doctoral thesis). Tlemcen: Abu Bekr Belkaid University.
- Djebbari, Z. (2014). *Self Confidence and Pronunciation Training To Enhance the EFL Speaking Competence: A Classroom Oriented Research on First-Year LMD Students at Abu Bekr Belkaid University, Tlemcen* (Doctoral thesis). Tlemcen: Abu Bekr Belkaid University.
- Hemche-Bereksi Reguig, H. (2015). *A Comparative Study of Teaching ESP in the LMD System in Algerian and French Universities: The Case of the Faculties of Sciences in Tlemcen and Paris-Sud Universities* (Doctoral thesis). Tlemcen: Abu Bekr Belkaid University.
- Khalidi, A. (2014). *Researching Learners' Difficulties and Strategy Use in Lecture Comprehension: The Case of First-Year EFL LMD Students Attending Lectures of Linguistics at Abu Bekr Belkaid University, Tlemcen* (Doctoral thesis). Tlemcen: Abu Bekr Belkaid University.
- Khelifi, K. (2014). *Learner and Teacher Readiness for Constructivism in the Algerian EFL Classroom: The Case of 3rd Year Literary Classrooms in Colonel Abd Elhadi Secondary School (Sidi Bel-Abbes)*. (Doctoral thesis). Tlemcen: Abu Bekr Belkaid University.
- Lee, S. (2010). *Current Practice of Classroom Speaking Assessment in Secondary Schools in South Korea*. Doctoral thesis. Australia: University of Queensland.
- Lira Gonzales, M. L. (2012). *A Teacher's Formative Assessment Perceptions and Practices in Oral Intermediate English Courses at the University of Montréal*. Montreal: Montreal University.
- Nadjia, Y. (2012). *An Assessment of Teaching the Speaking Skill: The Case of Second-Year LMD Physics Students at Abou Bekr Belkaid University-Tlemcen*. Doctoral thesis, University of Tlemcen.
- Özdemir-Çağatay, S. (2012). *Speaking Portfolios as an Alternative Way of Assessment in an EFL Context*. Ankara: Bilkent University
- Pršić, B. (2013). *Speaking Anxiety among High school EFL Students: Effects and Possible Remedies*. Master thesis. France: University of Niš.
- Rocío, S. A. (2012). *The Importance of Teaching Listening and Speaking Skills*. Spain: Complutense University of Madrid.
- SAK, G. (2008). *An Investigation of the Validity and Reliability of the Speaking Exam at a Turkish University*. Doctoral thesis. Ankara: Middle East Technical University.
- Torky, S. A. E. (2006). *The Effectiveness of a Task-Based Instruction Program in Developing the English Language Speaking Skills of Secondary Stage Students*. Egypt: Ain Shams University, Women's college.
- UsmaTrejos, L. M., Giraldo García, C., & Gómez, J. S. (2013). *Teaching speaking and speaking skill assessment in a 10th grade of a public high school from Pereira*.

Colombia: Technological University of Pereira Faculty of Fine Arts and Humanities Degree in English Pereira.

Wilson, J. T. S. (2006). *Anxiety in Learning English as a Foreign Language: It's Associations with Student Variables, with Overall Proficiency, and with Performance on an Oral Test: Doctoral thesis*. Editorial of the University of Granada.

Zidane, R. (2010). *Validating Holistic Scoring for the Assessment of EFL Learner's Writing Performance: The Case of Fourth Year Graduate Students at Abou Bakr Balkaid University Tlemcen* (Doctoral thesis). The University of Tlemcen

# APPENDICES

## APPENDICES

### APPENDIX 1

#### THE QUESTIONNAIRE ADDRESSED TO THE STUDENTS

*The current questionnaire aims at investigating the assessment of speaking skill among EFL learners. You are kindly invited to answer the below questions and thanked for your cooperation*

1\_Why did you choose to learn the English language?

- It is the global language
- I want to be an English teacher
- I want to go abroad
- Other

2\_In your opinion, what does it mean to learn a language?

- To be able to read writings in that language
- To be able to speak in that language
- To be able to listen to that language
- To be able to write in that language

3\_What is the skill you prefer to master more?

- The reading skill
- The writing skill
- The speaking skill
- The listening skill

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

4\_Are you satisfied with your level in speaking?

- Yes
- No
- Little bit

5\_ Are you satisfied with the way the speaking skill is taught?

- Yes
- No
- Little bit

6\_What are the tasks you prefer in an oral lecture?

- Group discussion
- Role plays
- Individual presentation
- Other .....

7\_What is the language you use to speak with your colleagues?

- English
- French

- Your dialect

8\_What is the language you use outside the classroom to speak with your teachers?

- English
- French
- Your dialect

9- Do you practice speaking with native speakers (in the net for example)

- Yes
- No
- Sometimes

10- What is more important in speaking English, in your opinion?

- Fluency
- Accuracy

11\_Do you like speaking tests?

- Yes
- No

Why:.....

12\_How do you feel when performing a speaking test?

- Comfortable
- Anxious

13 \_What are the difficulties you encounter in a speaking test?

- Lack of vocabulary
- Test anxiety
- Weakness in grammar
- The personality of the teacher

14\_If you are given the choice, which method you prefer in a speaking test?

- Face-to-face interview (i.e. An individual interview with your teacher)
- A paired interview (an interview with a colleague and the teacher observe)
- A group interview (for example doing a role play and the teacher observe)

15\_Are you satisfied with the way teachers rate your performance in speaking test?

- Yes
- No..... If No: what do you suggest as a better way of rating students' performance?

.....  
.....

16\_To which extent speaking tests can motivate students to learn more this skill?.....

.....  
.....

## **APPENDIX 2**

### **THE TEACHERS' INTERVIEW**

*The current interview aims at investigating the assessment of speaking skill among EFL learners. You are kindly invited to answer the following questions:*

- 1- For how many years have you been teaching speaking?
- 2- What are the tasks you usually prepare for a speaking lesson?
- 3- What are the difficulties you usually encounter when teaching speaking?
- 4- How do you proceed to assess your learners' speaking ability?
- 5- What are the difficulties you face when testing speaking?
- 6- What are the criteria you use to assess speaking?
- 7- What is more important in a speaking performance: fluency or accuracy?
- 8- What is more effective: formal or informal assessment?
- 9- How do you reach a reliable and valid speaking test?



**APPENDIX 3:**

**EUROPEAN LANGUAGE LEVELS - SELF-ASSESSMENT GRID**

© Council of Europe: Common European Framework of Reference for Languages (CEF)

		A1	A2	B1	B2	C1	C2
<b>Understanding</b>	<b>Listening</b>	I can understand familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.	I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear	I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect	I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort	I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided. I have some time to get familiar with the accent.
	<b>Reading</b>	I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues	I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.	I can understand texts that consist mainly of high frequency every day or job-related language. I can understand the description of events, feelings and wishes in personal letters	I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose	I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field	I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.

Speaking	Spoken Interaction	I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics	I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself.	I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events)	I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views	I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contribution skilfully to those of other speakers.	I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely. If I do have a problem I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it.
	Spoken Production	I can use simple phrases and sentences to describe where I live and people I know.	I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job.	I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions	I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	I can present clear, detailed descriptions of complex subjects integrating subthemes, developing particular points and rounding off with an appropriate conclusion.	I can present a clear, smoothly flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points
Writing	Writing	I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form	I can write short, simple notes and messages. I can write a very simple personal letter, for example thanking someone for something	I can write simple connected text on topics, which are familiar, or of personal interest. I can write personal letters describing experiences and impressions.	I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.	I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select a style appropriate to the reader in mind.	I can write clear, smoothly flowing text in an appropriate style. I can write complex letters, reports or articles, which present a case with an effective logical structure, which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.

## ملخص

إن التحدث باللغة الإنجليزية يعد من دون شك تحدياً لأي دارس للإنجليزية كلغة أجنبية. لكن التحدي الأكبر هو كيف يقيم الأساتذة قابلية الدارس للتحدث بهذه اللغة وعلى أي أساس يحكمون عليها بالقوة أو الضعف. إن الهدف من هذا البحث هو التحري عن الطريقة التي يتم بها تقييم مهارة التحدث لطلاب السنة الثانية ليسلنس في قسم الإنجليزية بجامعة أبي بكر بلقايد بتلمسان واقتراح حلول لرفع أداء الطلبة والأساتذة.

الكلمات المفتاحية: التحدث، التقييم، القابلية، الطلبة، الأساتذة، الممارسات.

## Résumé

Parler la langue anglaise est sans doute le défi de tout apprenant d'Anglais comme langue étrangère. Cependant, le plus grand défi est de savoir comment les enseignants évaluent-ils la capacité de parler d'un apprenant et sur quelle base la considère-t-on qu'elle est faible ou élevée. L'objectif de cette recherche est d'étudier l'évaluation de la compétence orale des étudiants du département d'Anglais à l'université de Tlemcen. Nous nous intéressons en particulier aux étudiants inscrits en deuxième année de licence, et ce, afin de remédier aux lacunes de l'expression mais aussi aux difficultés d'évaluation des enseignants par rapport à ces pratiques.

Mots-clés : parler, évaluation, capacité, enseignants, étudiants, pratiques.

## Summary

Speaking English language is without doubt the challenge of any EFL learner. However, the bigger challenge is how teachers assess the ability of speaking of a learner and on which ground they judge this ability as being high or low. The objective of this research work is to investigate the assessment of the speaking skill in the department of English at Abu-Bakr Belkaid University in Tlemcen, case of second year license students, and to propose some remedies to level up the students' ability to speak English and the teachers' assessment practices.

Key words: speaking, assessment, ability, teachers, students, practices.