

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC
RESEARCH



Tlemcen University
Faculty of Letters and Languages
Department of English
ESPT Lab

**AMBIGUITY OF THE WASHBACK PHENOMENON ON EFL
LEARNERS' ORAL PROFICIENCY: AN ACTION RESEARCH ON
SECOND YEAR ORAL TESTS AT TLEMCCEN UNIVERSITY**

*Dissertation Submitted to the department of English for the Fulfilment of the
Requirements for the Degree of Doctorate in Didactics & Assessment*

Presented by:

Ms. Houda DJEBBARI

Supervised by:

Dr. Nawel BENMOSTEFA

Board of Examiners:

Pr. BENMOUSSAT Smail	Prof. Tlemcen University	President
Dr. BENMOSTEFA Nawel	MC' A' Tlemcen University	Supervisor
Pr. MELOUK Mohamed	Prof. SBA University	External Examiner
Dr. BENSFAFA Abdelkader	Prof. SBA University	Internal Examiner
Dr. HAMZAOUI Chahrazed	MC' A' Ain Temouchent University Center	External Examiner

Academic Year: 2020/2021

Dedication

*To Whom I owe everything,
My Parents*

Acknowledgments

First and foremost, I would like to thank Allah for Health, willpower and energy to reach my goals,

I would like to thank my parents for the encouragements, care and continuous wise guidance,

I would never have been able to come to an end to my dissertation without the guidance of my highly esteemed supervisor Dr. Nawel BENMOSTEFA whose backup and support and advice are uncountable,

Special thanks go to the jury members.

Last but not least, I would like to thank all my teachers for all that they did for me and made me what I am today.

To all those I am truly grateful

Abstract

Successive attention of testing, measurement and evaluation research has been put on its content and on the residual effect tests may have on the educational process. However great the emphasis of testing within research may be, little is done concerning language testing practices and other parts still remain an enigma within the literature. Indeed, researchers often believe that there is a sense of ambiguity on the effects of testing on teaching; tests may lead either to success or failure depending on a number of changing variables. Thus, this research presents theoretical and practical frameworks to understand the relationship existing between oral tests, Washback and learners' overall oral proficiency. To this end, an action research was conducted on second-year LMD students at Tlemcen University. Besides, five research instruments were employed namely, learners' questionnaires, teachers' interview, oral proficiency tests, oral proficiency rating scale and classroom observation. In analyzing and interpreting data, the regression analysis, ANOVA, correlation significance analysis, and indirect proportion were the adopted statistical methods used. Results demonstrated that the nature of the study was a linear regression that specifies basically the relationship between the dependent variable (Y) to a combined function of the independent variable (X) and some unknown influencing factors (α). This denoted that there was a significant statistical association between the studied variables, the teaching method followed by the teacher and learners' oral proficiency, i.e., tests had a high impact on teaching and learners' oral proficiency. Eventually, a number of activities were proposed for a better teaching-learning-testing environment.

Table of Contents

Dedication.....II
Acknowledgments..... III
Abstract.....IV
Table of Contents.....V
List of Abbreviations and Acronyms.....X
List of Tables.....XI
List of Bar-Graphs and Figures.....XIII
List of DiagramsXIV
General Introduction..... 1

Chapter One: Background Account of the Selected Literature

1.1. INTRODUCTION..... 06
1.2. TEACHING AND TESTING..... 06
1.2.1. Testing in Education..... 07
1.2.2. Practical and Theoretical Considerations in Testing..... 08
1.2.3. Criteria and Principles of Testing..... 10
1.2.4. Proficiency in Testing: Means or Ends?..... 13
1.2.5. Course Design and Test Design: The Gap..... 15
1.3. WASHBACK PHENOMENON IN LANGUAGE TESTING..... 18
1.3.1. Washback Defined..... 19
1.3.2. Washback Dimensions..... 22
1.3.3. Washback: Positive, Negative, Neither or Both?..... 24

1.3.4. Teach-to-the-test Approach	26
1.3.5. The Washback Hypothesis.....	29
1.3.6. Previous Related Studies on Washback.....	32
1.3.7. A Critique of the Washback Studies.....	34
1.4. ORAL PROFICIENCY IN EFL CLASSROOM.....	35
1.4.1. The Notion of “Oral Proficiency”	35
1.4.2. Testing the Oral Proficiency	36
1.4.3. Problems in Testing the Oral Proficiency.....	39
1.5. INVESTIGATING THE AMBIGUITY OF WASHBACK PHENOMENON ON LEARNERS’ ORAL PROFICIENCY.....	40
1.6. CONCLUSION.....	42

**Chapter Two:
An Outlook into Research
Design and Methodology**

2.1. INTRODUCTION.....	45
2.2. LANGUAGE POLICY IN ALGERIA: REFORMS AND TESTING...	45
2.2.1. English Language Testing at the Tertiary Level.....	47
2.3. RESEARCH PHILOSOPHY AND PARADIGM.....	50
2.3.1. Research Roadmap, Aims and Research Questions.....	54
2.3.2. Research Timeline and Phases.....	55
2.4. SAMPLING ADMINISTRATION.....	55
2.4.1. Students’ Profile	56
2.4.2. Teachers’ Profile.....	56
2.5. RESEARCH INSTRUMENTS.....	56
2.5.1. Cycle one	57
2.5.1.1. Needs Assessment Questionnaire: Pre-Teaching Proficiency Test	57

2.5.1.1.1. One-Sample Kolmogorov-Smirnov Test.....	63
2.5.1.2. Washback Questionnaire.....	64
2.5.1.3. Learners’ Oral proficiency Test.....	68
2.5.1.4. Oral Proficiency Rating Scale.....	70
2.5.1.5. Teachers’ Interview.....	71
2.5.1.6. Classroom Observation Phase I.....	74
2.5.2. Circle Two.....	75
2.5.2.1. Washback Questionnaire II.....	76
2.5.2.2. Classroom Observation Phase II.....	78
2.5.2.3. Oral Proficiency Test.....	79
2.5.2.4. Oral Proficiency Rating Scale.....	80
2.5.2.5. Post-Test Questionnaire.....	81
2.6. Conclusion.....	84

**Chapter Three:
Raw Data Processing**

3.1. Introduction.....	86
3.2. Data Analysis Procedure.....	86
3.2.1. Qualitative Data Analysis.....	87
3.2.2. Quantitative Data Analysis - SPSS –.....	88
3.3. CYCLE ONE DATA ANALYSIS.....	88
3.3.1. Needs Assessment Questionnaire Analysis.....	89
3.3.2. Washback Questionnaire Analysis.....	94
3.3.3. Learners’ Oral Proficiency Test Results.....	96
3.3.4. Teachers’ Interview Results.....	102
3.3.5. Observation Phase I Analysis.....	109
3.3.6. Oral Proficiency Rating Analysis.....	110

3.4. CYCLE TWO DATA ANALYSIS.....	114
3.4.1. Washback Questionnaire Analysis.....	114
3.4.2. Observation Analysis Phase II.....	117
3.4.3. Oral Proficiency Test Results.....	118
3.4.4. Proficiency Rating Scale.....	119
3.4.5. Post Test Questionnaire Analysis.....	122
3.5. INTERPRETATIONS OF THE MAIN RESULTS.....	133
3.6. CONCLUSION.....	138

Chapter Four
Considerations and
Implications for Innovation

4.1. INTRODUCTION.....	140
4.2. NEW PARADIGMS FOR TESTING: <i>Call for Urgent Progress</i>	140
4.3. IMPLICATIONS FOR CHANGE: <i>Teachers as Professional Assessors</i>	143
4.4. LAB-BASED ASSESSMENT: <i>Technology Introduction</i>	149
4.4.1. Testing Speaking Reconsidered.....	151
4.4.2. Online Assessment in Oral Testing: Fine Grading Techniques.....	152
4.5. Introducing Online Blended Teaching In Time of Covid-19 Pandemic	161
4.6. FACTORS INFLUENCING LANGUAGE ASSESSMENT.....	167
4.6.1. Test Anxiety as a Barrier in Testing.....	168
4.6.2. Factors Affecting Test Anxiety.....	171
4.7. NEW DIRECTIONS TO WASHBACK PHENOMENON.....	174
4.8. CONCLUSION.....	175

GENERAL CONCLUSION.....	176
BIBLIOGRAPHY.....	180
Appendix A : Needs Assessment Questionnaire.....	199
Appendix B : Washback Questionnaire.....	202
Appendix C : Learners’ Oral proficiency Test	204
Appendix D : Teachers’ interview.....	206
Appendix E : Observation for the teacher	210
Appendix F : Oral Proficiency Rating Scale.....	212
Appendix G : Washback Questionnaire.....	214
Appendix I : Oral proficiency Test.....	216
Appendix J :Post-test Questionnaire.....	218
Appendix K : Proficiency rating scale II.....	222
Appendix L : Students Scores.....	224
Appendix M : SPSS Tables.....	226
Appendix N : Correlation Tables.....	229
Appendix O : SPSS Output.....	231
Appendix P : SPSS Certificate.....	234
Appendix Q :Speechace Invoice.....	236

List of Abbreviations and Accronyms

1S.M: One Sample Median

2I Tt: Two independent t-Test

A.F: Absolute Frequency.

A.R: Action Research

ACTEFL: American Council on the Teaching of Foreign Languages

ANOVA: Analysis of Variance,

BT: Binominal Tests

EFL: English as a Foreign Language

ELPT: English Language Proficiency Test

ESL : English as a Second Language

ETS Education Testing service

FdT : Friedman Test

FT : Ficher's Exact Test

K.S : Kruskal Wallis

KS: kolmogorov smirnov

LMD: Licence Master Doctorat

OPI: Oral Proficiency Interview

Q2: Chi-Square Test

r: Pearson Coefficient Correlation

R.F: Relative Frequency

SD: Standard Deviation

SPSS: Statistical Package for Social Sciences

SS: Sums of Squares

TEFL: Teaching English as a Foreign Language

RF: Relative Frequency

SD: Stadarnd Deviation **W.M :** Wilcoxon-Mann Whitney test

SRRS: Speech Recognition Rater System

List of Tables

Table.1.1. Types of Validity.....	12
Table1.2. Principles of Language Tests (Adopted from Bailey,1998).....	13
Table1.3. Influence of Tests on Teaching Materials	18
Table1.4. Selected Definitions of Washback.....	21
Table.1.5 Washbach Dimensions.....	24
Table 1.6. Connection between Hughes’ Trichotomy and Alderson and Wall Hypothesis (1993).....	31
Table1.7. Early Studies on Washback.....	34
Table1.8. Assessing Fluency(adopted from Towell,2006).....	38
Table 1.9. Problems of Oral Proficiency Assessment.....	40
Table 2.1. LMD System Restrictions.....	49
Table 2.2. Needs Assessment Questionnaire Objectives.....	60
Table 2.3. Questionnaire Objectives.....	61
Table 2.4. Coding strategy.....	61
Table 2.5. Internal consistency.....	62
Table 2.6. Squared Multiple Correlation.....	63
Table 2.7. Questionnaire objectives.....	65
Table 2.8. Kendall's Test norms.....	67
Table 2.9. Oral Proficiency Test Objectives.....	69
Table 2.10. Assessment Grid.....	70
Table 2.11. Interview Types(Bhat, 2015).....	72
Table 2.12. Teachers’ Interview Objectives.....	74
Table2.13. Course Design.....	75
Table 2.14. Questionnaire Objectives.....	77
Table 2.15. ACTEFL Proficiency Assessment.....	79
Table 2.16. ACTEFL proficiency assessment Stages.....	80

Table 2.17. ACTEFL Proficiency Assessment Grid.....	81
Table 2.18. Post Test Questionnaire Objectives.....	82
Table 2.19. Post Test Questionnaire Pre-Tests.....	84
Table 3.1. Speaking Problems faced by Learners.....	89
Table 3.2. Teachers' Washback Questionnaire I	94
Table 3.3. Internal Consistency	97
Table 3.4. Observation Grid Results.....	109
Table 3.5. Internal Consistency estimations.....	110
Table 3.6. Teachers' Washback Questionnaire II.....	115
Table 3.7. Observation II Grid Results.....	117
Table 3.8. Assessment Grid.....	118
Table 3.9 Correlation Significance	136
Table 3.10. B Constant Coefficient	137
Table 4.1. Different stages of Role play.....	148
Table 4.2. Fluency Assessment Criteria adopted from Speechace.com	159
Table 4.3. Scoring Criteria (adopted from Speechace.com)	159
Table 4.4. Online Teaching Vs Face to face Classroom.....	165
Table 4.5. Strategies to Deal with Learning in times of Pandemic.....	167
Table 4.6. Test Anxiety Symptoms (Adopted from Horwitz,2002)	169

List of Bar-Graphs and Figures

Bar-Graph 3.1. Learners’ Oral Proficiency Component Awareness.....	90
Bar-Graph 3.2. Human rater scores.....	98
Bar-Graph 3.3. Automatic Rater Scores.....	100
Bar-Graph 3.4. Automatic Rater Scores Curve.....	100
Bar-Graph 3.5. Teachers’ Beliefs about Learners’ Oral Proficiency Level.....	103
Graph 3.6. Teacher’s Most Challenging Phases.....	105
Bar-Graph 3.7. Teacher’s Assessment Criteria.....	107
Bar-Graph 3.8. Test Scores.....	119
Figure1.1. Models of Washback (Bailey,1994:264).....	32
Figure 2.1. Cyclical Model AR Model(Kemmis and McTaggart, 1986).....	52
Figure 2.2. Needs Analysis Model (adopted from Dudley-Evans (1998).....	61
Figure 4.1. Advantages of Role Play Adapted From Chesler And Fox (1966)..	148
Figure 4.2. Speechace Platform.....	153
Figures 4.3. Speech Recognition Page.....	154
Figure 4.4. Speechace Pronunciation Offers.....	154
Figure 4.5. Speechace Speech Activities	155
Figure 4.6. Speechaace Word Pronunciation Topics.....	155
Figure 4.7. Speechace Sentence Pronunciation.....	156
Figure 4.8. Speechace Fluency Assessment.....	156
Figure 4.9. Speechace Fluency Worksheets.....	157
Figure 4.10. API Assessment Guide.....	160
Figure 4.11. Students IELTS Scoring.....	160
Figure 4.12. Pearson.com.....	165

List of Diagrams

Diagram1.1. Factors influencing the Washback (Adopted from Shohamy et al, 1996).....	22
Diagram 1.2. Bailey’s Trichotomy.....	30
Diagram 2.1. Cyclical AR Model (Kemmis and McTaggart,1986)	51
Diagram 2.2. The Process of Action Research.....	52
Diagram 2.3. Action Research Plan.....	53
Diagram 2.4. Action Research Cycles.....	53
Diagram.2.5. Reasearch Variables.....	55
Diagram 2.6. Research Phases.....	55
Diagram 2.7. Cycle One Research Instruments	57
Diagram 2.8. Cycle Two Research instruments	57
Diagram 2.9. Needs Analysis Model.....	59
Diagram 2.8. Oral Proficiency Assessment.....	65
Diagram3.1 Indirect Proportion between Variables.....	138
Diagram 4.1. Speechace Assessment Criteria (adopted from speechace.com) ...	158
Diagram 4.2. Foreign Language Anxiety Adopted from Horwitz(1986).....	170



General Introduction

General Introduction

Language researchers appear to steadily acknowledge that the field of language testing seems to be a relatively new trend that has entered the field of applied linguistics and didactics. Thus, it seems impractical to tackle the challenges of teaching without considering the demanding requirements of testing. In fact, testing field is in vogue within recent research worldwide and the introduction of new testing system is the responsibility of teachers and policy makers alike. A glimpse through the history of language testing provides a fascinating picture of the varied testing systems that have ever been introduced to this field.

The educational system is flooded with a variety of high stakes testing, this renders testing and assessment increasingly questioned and investigated. Assessment affects our learners' careers and future decisions. In fact, considering the existing efforts regarding assessment development, there has been very slow movement in the renewal of new ideas meant to change current practices. Looking deeply at the content and approaches used in assessment practices in higher education, one may diagnose a profound gap between delivering knowledge and assessing it. Major focus is placed on students' grading and scoring, neglecting at a certain extent the processes of teaching, learning and assessing how students will learn after feedback. In other words, assessment is not sufficiently well designed, equipped and tracked.

Washback is a remarkable concept referring to the influence of testing on teaching/ learning. In fact, the widespread use of examinations at the level of all our educational, social and personal lives has rendered washback a challenging, worth researching trend. The appealing influence of tests on teaching has always been considered a brightening spot of research since it influences the way teachers teach, select materials, design courses, and test.

In this challenging global age, and owing to the ongoing globalization as a pervasive phenomenon, a sound attention should be driven towards the importance of developing learners' oral proficiency. In fact, developing learners' speaking abilities is

General Introduction

a worth discussing topic within the areas of curriculum, instruction, and assessment. In a more particular image, the feasibility of assessing learners' oral proficiency has put teachers in a maze of choices and challenges. Speaking is often regarded as one of the most intricate skill to test, since it involves a mixture of all language skills that may have no correspondence with each other. (Kitao & Kitao: 1996).

In the quest of investigating what may be the causes of the low achievement rates within university students; this research work will be proposed aiming at finding the possible relationships between testing learner's oral proficiency, washback phenomenon and learners' test scores. In other terms, the present research work is intended to investigate the effects tests have on both the teaching of speaking and the overall learners' speaking proficiency. This research is carried out to rise teachers' awareness of the washback phenomenon, and will accompany learners throughout the term with facilitating strategies that will help them develop their speaking proficiency.

This research work attempts to demonstrate the ambiguity of washback phenomenon on the teaching strategies adopted by the teachers to develop learners' oral proficiency. It also seeks to measure the correlation between teaching-to-the-test, washback and learners' oral proficiency. To sum up, the main objectives are:

- Analyze the influence of testing on teaching and its effects on learners' general oral proficiency,
- explore the influence of tests on learners' general oral proficiency,
- examine learners' and teachers' perceptions of washback effect,
- compare learners' speaking performance while applying the teach-to-the-test approach during the course of instruction and while following an ordinary teaching approach,
- seek teachers awareness of course design and test design.

Hence, the study raises the subsequent research questions:

- How may negative/positive washback influence learners' oral proficiency?
- What is the relationship between learners' scores and the teaching method?
- What is the relationship between scores and oral proficiency?

Based on the raised questions, these hypotheses are suggested:

General Introduction

- Negative washback limits learners' oral proficiency and positive washback help improve learners' oral proficiency and develop their general speaking competence.
- Teaching methodology followed by the teacher influences learners' scores.
- There exist a positive statistical correlation between learners' oral proficiency and tests scores.

To achieve satisfactory answers to the asked questions, this research will be divided into four chapters. The first one is devoted to a relevant background account of the literature surrounding the topic of the research work. It will introduce key-concepts definitions, including some previous studies conducted within the same area of research.

The second chapter is based on a descriptive approach; it turns around the description of the research setting and the sampling procedure, research design including the description of the applied cycles of action research and the instruments used for data collection. A triangulation process will be used to collect data both qualitatively and quantitatively, using questionnaires, interviews, oral tests, classroom observation and a proficiency rating scale.

To adequately investigate the collected data from chapter two, chapter three ventures to analyze data both qualitatively and quantitatively hoping to find out answers the research questions settled. Statistical methods of data analysis will be applied using SPSS version 22 to reinforce the reliability of the results achieved.

Finally the fourth chapter presents some practical considerations and suggestions to develop our learners' oral proficiency and avoid negative washback in our classes. It also offers some up-to-date techniques of test design, using online software that may support teachers in their assessment procedure.

Chapter One: Background Account of the Selected Literature

1.1. INTRODUCTION

1.2. TEACHING AND TESTING

1.2.1. Testing in Education

1.2.2. Practical and Theoretical Considerations in Testing (criteria, principles)

1.2.3. Criteria and Principles of Testing

1.2.4. Proficiency in Testing: Means or Ends?

1.2.5. Course Design and Test Design: The Gap

1.3. WASHBACK PHENOMENON IN LANGUAGE TESTING

1.3.1. Washback Defined

1.3.2. Washback: Positive, Negative, Neither or Both?

1.3.3. Teach-to-the-test Approach

1.3.4. The Washback Hypothesis

1.3.5. Previous Related Studies on Washback

1.3.6. A Critique of the Washback Studies

1.4. ORAL PROFICIENCY IN EFL CLASSROOM

1.4.1. The Notion of “Oral Proficiency”

1.4.2. Testing the Oral Proficiency

1.4.3. Problems in Testing the Oral proficiency

1.5. THE AMBIGUITY OF WASHBACK PHENOMENON ON LEARNERS’ ORAL PROFICIENCY

1.6. CONCLUSION

1.1. INTRODUCTION

Diving into the history of language testing, one would suppose that language researchers seem to be concerned mostly with the theoretical approaches and methods neglecting, at a certain extent, their practical content within a rational milieu. In fact, in educational literature, testing seems to be an essential paradigm in language assessment. Its weight is mirrored through the reflects of the available literature devoted to this well-known facet of education. The growing interest of testing within teaching driven by the power of language tests and high stakes exams in particular, has created a line of arguments among researchers: a phenomenon known as “Test Washback”. The reservation of washback phenomenon policies in education seems to be among the most fundamental issues occupying language research.

Thus, the present chapter has a two folds aim: first, it seeks to review the existing literature related to the study and highlights concepts like test impact, curriculum alignment, and oral proficiency. Second, it aspires at bridging the gap between the influence language tests have on learners’ overall performance in general and their oral proficiency in particular.

1.2. TEACHING AND TESTING

Looking back at the past century of the teaching-testing methods and techniques would give us a portrait of the varied explanations on how to teach and test a language. In fact, it is generally known that teaching and testing are two sides of the same coin, that there is neither teaching without testing nor testing without teaching (Benmoussat, 2003).

In formal classroom practices, tests remain an unavoidable part in language teaching because there is a perceived need for a method to measure learners’ language ability (Brown, 2004). Although testing is of an axiomatic function, tests can be closely associated with pedagogical purposes (Bachman and Palmer, 1996). Thus, teachers, state policy makers and test developers seek ways to drive pedagogical benefits from both teaching and testing (Watanabe, 2004). Language tests are usually used to provide learners with adequate feedback about their growth and teachers’ classroom practices. In this sense, Flavell (1981: 1) states that: *‘A test is seen as a natural extension of classroom work, providing teacher and student with useful*

Chapter One: Background Account of the Selected Literature

information that can serve each as a basis for improvement'. Put differently, test contents are put as tools for classroom activities improvements, feedback about learners' progress or failure and as a remedy for some teaching practices.

1.2.1. Testing in Education

Retrospectively, tests have always been used as a means of control over time (Hughes, 1986). Tests are used by policymakers to manipulate stakeholders in many parts of the world, either by controlling the educational system or imposing new teaching methods (Cheng, 2008). In this context, it is always posited that testing and assessment are "*the darling of the policy-makers*" (Madaus, 1988). Testing is viewed as one of the chief agents for making change in the educational systems; however, it is always put under criticism and pressure. In this sense, Petrie (1997:175) concludes his study by stating that '*it would not be too much of an exaggeration to say that evaluation and testing have become the engine for implementing educational policy*'.

Later, testing has always been an interesting area among teachers and researchers alike. Standardized tests are a powerful tool used by the government to improve educational quality at the national level. In this sense, Herman (1992:2) believes that '*testing advocates and many policymakers still view testing as a significant, positive, and cost effective tool in educational improvement*'.

Before diving into testing challenges in education, it is worth defining the concept as put by scholars. In fact, it is believed that testing is a device or an instrument that measures learners' general competences and the linguistic and grammatical competence in particular. According to Carrol:

The purpose of testing is always to render information to aid in making intelligent decisions about possible courses of action. Sometimes these decisions affect only the future design or used of the tests themselves, in which case we are dealing with solely experimental uses of tests. Sometimes the decisions have to do with the retention or alteration of courses of training, as when one decides that poor tests results are due to in effective training. Carrol (1965: 364):

Chapter One: Background Account of the Selected Literature

Teachers routinely adopt tests to detect learners' potentials and weaknesses and evaluate their progress and achievements (Alipanahi, 2016). Tests are commonly viewed as a means to measure innovation, guide and direct the curriculum (Alderson, 2004).

However great the emphasis of testing within research may be, little is done concerning language testing practices and some parts still remain an enigma within the literature. In fact, an existing ambiguity about the influence tests may have on teaching has been an interesting scope of research, since it is believed that in certain contexts, the same test may have various results under different circumstances and teachers may use very diverse strategies to prepare students for tests (William, et. al, 2004). Therefore, teachers must have a sound knowledge about *how to* design and respond to new testing systems, *what* reasons should accompany those tests, and *what* educators need to know to make sure that testing helps learners grow and develop rather than keeps them becoming slaves to tests (Kaufman, 2004).

It is, thus, indispensable for all the educational community to cooperate, assist and engage to well recognize and assess the possible influences of the use of testing on all the interrelated facets of the educational systems.

1.2.2. Practical and Theoretical Considerations in Testing

Language teachers routinely encounter within their daily experience pervasive obstacles that interfere with having suitable learning environment. Thus, awareness must be driven on the new testing frames, roles of classroom testing, testing principles and what makes a test qualified as a good test. While some researchers view tests as having negative consequences *per se* for teaching and syllabus design (Wiseman, 1961), others view that tests are positive as potential instruments for educational reform (Pearson, 1988). In this line of thought, Valette (1977) describes that classroom assessment accomplishes three actions namely;

- Definition of course objectives.
- Tests stimulate student progress.
- Evaluation of classroom achievement.

Chapter One: Background Account of the Selected Literature

- *Definition of Course Objectives*

Valette (1977) believes that before engaging in teaching in general and in course design in particular, teachers must clearly envision the course objectives. Therefore, teachers should systematically set short term course goals as well as items to be taught within the formal classroom setting. By doing so, teachers will be sure that their teaching will be rationally-oriented and that tests will show how close each student has attained the settled objectives.

- *Tests Stimulate Students' Progress*

Valette (ibid) explains that tests should give a clear picture of how well students handle specific element of the target language. Thus, learners themselves are expected to demonstrate their performance errors. Considering this view, he (1977:4) states that *'The test best fulfills its function as part of the learning process if correction performance is immediately confirmed and errors are pointed out'*.

- *Test Evaluates Class Achievement*

Valette (ibid) also believes that through current testing, teachers can have an idea and can decide which features of the course are difficult for individuals, which parts had been well perceived by students. Teachers then, can decide on the content of the extra remedy sessions and how best they may assist each learner in analyzing the mistakes made on a given tests. Testing also helps teachers discover which classroom objectives have been met, evaluate the effectiveness of new teaching methods, different approaches and new materials. In a word, testing should be viewed as a *bridge-building* practice between teaching, learning and the current classroom tests as mirrors in which teachers and students see their reflections clearly Valette (1977) cited in Benmostefa, (2013:70)

1.2.3. Criteria and Principles of Testing

Generally speaking, there are myriad of different language tests, they may range from school exams, university entrance exams or low -high stakes exams. They aim at providing the test taker with some sort of standardized qualification. Language tests

Chapter One: Background Account of the Selected Literature

provide a reasonable, standardized assessment as a source for good qualifications. They are then, concerned with a restricted classroom environment, following a set of rules and principles. In view of this, one can differentiate two types of tests: *proficiency tests* and *achievement tests*. The former deals with the amount to which the test taker has reached the course goals and objectives, therefore, proficiency. Accordingly, they prepare teachers for remedy sessions and allow further interpretations for future performances. However, achievement tests usually pursue the principles of designing the test as the course was taught (Vollmer, 2000).

In fact, developing a language test that is qualified to account for different language aspects is not easy, and is deemed a challenge faced by teachers. In this context, Brown (1994:253) claims that “*one of the main biggest obstacles to overcome in constructing adequate tests is to measure the criterion and not inadvertently something else*”. As such, he puts forward three self-evident requirements to certify a test as “*good*”, notably: practicality, reliability and validity as explained bellow:

- *Practicality*

Practicality is in fact, an integral part of the usefulness of language tests that affects many aspects of the test. It is seen as the extent to which a test is practicable enough in terms of resources, settings and ease of administration. A test is said to be practical if some considerations are taken into account. For instance a test is practical when it is not extremely costly, stays within suitable time constraint, moderately easy to administer and eventually has a scoring process that is precise and time efficient. For instance, a test that is expensive is and that takes five to six hours to be completed is not practical (Brown; 1994).

- *Reliability*

Reliability designates the extent to which a test gives trustworthy results. A reliable test is consistent and steady, in the sense that the same test is given to the same students on diverse settings and yields similar results. Test takers should get similar scores on diverse test formats. Accordingly, the same test may encounter to measure the same language variables using the same methods of testing. Herein, Harmer (2001:322) puts that, “*In practice, “reliability” is enhanced by making the test*

Chapter One: Background Account of the Selected Literature

instructions absolutely clear, restricting the scope for variety in the answers, and making sure that the test conditions remain constant". Thus, reliability is of paramount importance in deciding whether or not the test gives consistent results.

- *Validity*

Validity shows the extents to which a test determines what is intended to measure. Garrett (1964:30) states that "*A test is valid when it measures what it claims to measure.*". On his side, Abbott (1992:178) claims that "*A test cannot be a good test unless it is valid. The essence of validity means the accuracy with which a set of test scores measures what it claims to measure.*" Therefore, a valid exam is a means to which the test maker gives consistent measures and clear valid judgments. For instance, a ruler is a valid instrument but impractical to measure the distance from Tlemcen to Algiers, even though it is reliable enough in general but not valid sufficiently in such a condition.

Alderson (1975) believes that validity contains a set of criteria among which, content validity, indicates the degree to which the test used measures what it intended to measure. In this sense, to be qualified as a valid test, the contents should constitute the language skills and structures required. In other words, Benmostefa (2014:53) states that "*the aspect of validity is based on the degree to which a test adequately and sufficiently measures the particular skills it sets out to measure, what is called content specification, in other words, the extent to which the content of the test matches the instructional objective*".

At this level, reading the literature surrounding the criteria of what constitutes a valid test; one would find a maze of theories different strategies and methodologies proliferating. For instance, Cronbach and Meehl(1955) classify four types of validity among which: predictive validity, concurrent validity, content validity, and construct validity. The first two types of validity are classified as "criterion-oriented-validation" because they predict mainly when the test is administered. The following table summarizes all the types as follows:

Types of Validity	Interpretation
Predictive Validity	is studied when the criterion get some time after the administration of the test
Concurrent Validity	is inspected when the test score is determined, it can be analysed when the test is planned as a subsequent for 2 nd test.
Content Validity	it seeks to examine if the test is appropriate and suits the content and the settings of a test. As Xi (2008:178) beleives, <i>“A direct language test has to show face or content validity by demonstrating its resemblance of “real life” language situations in the setting and linguistic content.”</i>
Construct Validity	is invested when the test maker wants to measure if the test reflects the theory or the fundamentals of the test.

Table.1.1. Types of Validity

Therefore, considering these criteria turns the test at any level, a more complicated, challenging phenomenon. Thus, language tests should be designed by teachers who have received previously an adequate training of test design or by curriculum developers or by language testing specialists in general.

Designing an adequate language test appears to be a daunting task that requires some efforts from the part of the teacher. To achieve at a large extent a satisfying image, Bailey (1998) suggests four main principles to reach a ‘good’ test. These principles are put in the following table:

“Know what to test”	“Content Concentration”	“Be Ready for the Test”	“Positive Wash-back”
<ul style="list-style-type: none"> • Test makers must know what is expected to be measured and that test designers must expect how the 	<ul style="list-style-type: none"> • The content of the test should be based on the courses already administered, should also suit students age 	<ul style="list-style-type: none"> • Providing appropriate review that suits both good students and under achievers Swain(1984) 	<ul style="list-style-type: none"> • in order to achieve it teachers must be provided with clear scoring criteria course objectives and

test takers will answer.	interest and preferences.		test content
--------------------------	---------------------------	--	--------------

Table 1.2. Principles of Language Tests (Adopted from Bailey, 1998)

1.2.4. Proficiency in Testing: A Means or an End?

Within an arena of high-stress and high-stakes testing, teachers are always ahead of a huge pressure to promote and develop their students' scores. This pressure has led teachers, parents, stake holders to offer classroom instruction that intends to develop only the needed items on high-stakes tests, featuring "*clone items*" testing rather than '*mastery-based*' testing. Hence, the question that it worth raising here is that are our teachers testing for proficiency of testing for grades?

Tests help learners self-assess their levels, it also helps teachers evaluate the teaching method followed and how much of classroom instructions have been learned. However, testing may do harm to the teaching learning process.

In fact, testing for grades not only distorts tests validity, but also offers language learners with the short end of the educational stick, i.e., both teachers and learners will end up with creatures controlled by grades. This way will decrease learners' ambitions and stop their creativity, minimizing mastery-based testing. Considering this performance as being unethical, it deprives learners from broad understanding of concepts and ideas.

In a recent strategy applied by a number of Chinese schools, a policy of a 'New Kind of Classrooms: No Grades, No Failing, No Hurry' was adopted to stop the stress of learning for grades and raise learners' general competences. This policy joins the idea proposed by Chomsky (ibid) who believes in the danger of assessing learners. He seemingly calls for stooping the process of assessing and grading learners especially during early education. He, in this sense, assumes that:

You don't have to assess people all the time... People don't have to be ranked in terms of some artificial [standards]... So you are giving

some kind of a rank, but it's a rank that's mostly meaningless. And the very ranking itself is harmful. It's turning us into individuals who devote our lives to achieving a rank. Not into doing things that are valuable and important.

Reflecting on the above ideas, and contextualizing the issue, one may presuppose that teaching-to-the-test applied in high-stakes exams may have the same results stated earlier. Learners will become slaves to items proposed in the curriculum and cannot broaden their knowledge. Within the same line of thought, teachers will no more become victims of a *score-boosting* game that they cannot and will never win (Chomsky:2015).

Hence, if policy makers exploit tests with clearly set instructional targets, and well defined goals, teachers can center their efforts getting learners to master what they're supposed to learn and grow rather than following item-based testing which leads to limiting learners' potentials Chomsky (ibid). Besides, whether tests will result positive or negative outcomes, other dimensions should be considered, such as *who* is conducting the test design process, *where* it is conducted, *when* assessment practice takes place, in addition to *why* to assess (the rationale), and *how* teachers used the different approaches within the educational context.

1.2.5. Course Design and Test Design: The Gap

Researches surrounding washback phenomenon have demonstrated that exams, be them high stakes or low stakes, take place at different levels and diverse degrees in the teaching/learning. The relationship between testing and educational practices represents "*the ripple effect*" to use Barnes' term (2000:632). Consequently, it is generally acknowledged that exams have a great influence on teachers' practices, attitudes and evaluation strategies. Arguably, Odo states that:

Researchers are becoming progressively more aware of the negative social impact large-scale high-stakes tests can have on the lives of learners – particularly those who are most vulnerable – when the results of these tests are used to make

Chapter One: Background Account of the Selected Literature

decisions that unfairly limit the life choices of these learners

Odo(2012:2)

The testing literature reveals that examinations, especially high stakes exams tend to force instructors to follow the teach-to-the-test approach as a means of instruction. Cheng (2005) conducted a study in her Hong Kong Certificate of Education Examination (HKCEE) research work, she concluded that teachers were concerned about the ways their students would pass the revised test. One teacher declared that she would feel guilty if her students would not be familiarized with the test layout and content before conducting the test.

On his side, Tsagri(2009) researched the influence of tests on participants' perception and on teachers' material design. Results showed that teachers were forced and tried to cover all the materials on the prescribed syllabus. Notwithstanding the harm that might be caused by high stakes exams, it has been noticed that this might have a positive effect on the teaching learning process. It will render teachers more accountable, more encouraged and more caring and spend more time on meaningful learning tasks. Conversely, tests may have negative consequences that may have positive effects on learners' psychological state and on teachers' instructions. In this sense, Gregory and Burg conclude that:

The extent to which a teacher provides explicit structure during lessons such as providing frequent previews and reviews, and reduces the density of instruction and content input have both been identified as potentially reducing the debilitating effects of test anxiety on student achievement. Burg (2006:44)

Teachers are the key component in the introduction of reforms at the educational system, they are the “*front-line*” mediums for the washback process related to instruction Bailey (1999).

Methodology washback implicates how instructional practices may be adjusted through high stakes examination reconstruction, i.e., teaching practices, methods and techniques employed by the teacher, classroom assessment practices, formative

Chapter One: Background Account of the Selected Literature

assessment and eventually summative assessment. It has been generally agreed on that teachers play a pivotal position in promoting different structures of washback, as believed by Spratt who states that:

they play a significant role in determining the types and degrees of washback impact, promoting (or inhibiting) positive washback. Keeping this in mind, teachers (and inspectors—expert teachers) were targeted as the main population of the study. Spratt (2005:34)

Herein, teachers generally use materials, courses, textbooks, and items in the syllabus to deliver instruction. In this vein, Wall (2012:79) significantly views that the influence washback has on teaching materials can be visualized when both teachers and students “pay more attention to certain parts of the teaching syllabus at the expense of other parts because they believe these will be emphasized on the test”. This latter was inspired by early studies of washback which proved that teachers rely on contents that surround the test in designing their teaching materials. This action is called *curriculum alignment* for example (Madaus,1988 ,Wall and Alderson (1993:126) scrutinize that: “the new exam has had a demonstrable effect on the content of language lessons”.

Teachers influenced by language assessment tend to create their own materials and alternate to previous examination samples. However, teachers tend to limit teaching materials to meet the test requirements. Results from the previous studies about the influence tests have on teaching materials are summarized in the following table:

Researchers	Findings
Andrews (1995:80) and Lam (1995:35)	<i>"positive washback is evidenced by teachers creating more authentic materials from the mass media, [and] producing meaningful learning activities"</i>

Chapter One: Background Account of the Selected Literature

Andrews (1995:80)	speaks of the role played by published materials, reporting that teachers spent an <i>“estimated two-thirds” of classroom [instructional] time on exam-related published materials, which, as the author advocates, “represent a limiting of focus for teachers and students rather than a broadening of horizon”</i> (80).
Watanabe (2000:22)	<i>"tried to innovate during exam preparation classes ... using a variety of self-made material"</i>
Gorsuch (2000 :55)	<i>although the reform urged the use of the all four language skills equally, the examination administered to the same schools emphasized testing the knowledge of vocabulary, grammar and language usage.</i>
Agrawal (2004 :23),	<i>while the teaching syllabus focused on developing oral skills, instructors tended to marginalize these skills on their teaching agenda because they were not included in the exams written by the authority of education</i>

Table.1.3. Influence of Tests on Teaching Materials (adopted)

For Alderson and Wall(1993:127): *“we need to look closely at classroom events [by using direct research methods for data collection] in particular, in order to see whether what teachers and learners say they do is reflected in their behavior”*. Thus, the findings might not be reliable enough since they relied mainly on questionnaire analyses which have proved inadequate evidence for making clear the effects of on examinations.

According to Shohamy (1992) this gap between test design and course design is the result of an external authority that impacts test takers, teachers and stakeholders as a whole. She (1992:299) states:

The power and authority of tests enable policy-makers to use them as effective tools for controlling educational systems and

prescribing the behavior of those who are affected by their results administrators, teachers and students. Schoolwide exams are used by principals and administrators to enforce learning, while in classrooms, tests and quizzes are used by teachers to impose discipline and to motivate learning.

Shohamy(1992:299)

Notwithstanding the variation in views, one may accept it as a truth that if the test is well designed and appropriately employed by teachers, there might be no gap between what is designed and what is expected from learners, this will create a positive attitude and strong inspiration to learn and grow, on the one hand and also help instructors collect feedback and better their teaching/testing practices. However, if poorly designed, the test will result poor outcomes and direct teaching and learning off track. This influence is called washback.

1.3. WACHBACK PHENOMENON IN LANGUAGE TESTING

During the past decade, successive attention of testing, measurement and evaluation research has been put on its content and on the residual effect tests may have on the educational process. Many scholars and policy makers in language testing sphere have driven great attention on the influence tests, especially high stakes, may have on the teaching/learning process. Under the stress to help students reach satisfactory results on such tests, teachers and administrators tend to give more importance on test content and the format devoting more time in test preparation, in other terms, teaching exam-related content or what is called in the testing literature item-teaching rather than teaching different linguistic variables. It is generally acknowledged that in this specific context, tests apply a powerful influence on teachers and learners; this phenomenon has been labeled as being the washback or backwash effect (Alderson and Wall, 1993). Considering its importance, the following section is devoted to analyze this phenomenon and its related issues.

1.3.1. Washback Defined

Looking through the amount of the teaching and testing stockpile of knowledge exposes that despite the plethora of empirical evidence about terminology, there seem to be a variation in researchers' perception of the conceptions of the washback phenomenon. In this line of thought, several definitions of the concept have been offered in the testing literature as believed by Bailey (1999: 3) : “*definitions of washback are nearly as numerous as the people who wrote about it*”. Shohamy, et.al, (1996:298) defines it very plainly: “*the connections between testing and learning*”; whereas Gates (1995:101) perceives it as “*the influence of testing on teaching and learning*”. Thus, the way in which high stakes exams affect largely the teaching process is called the back-wash or the wash-back phenomenon, test impact, curriculum alignment and test feedback. In this sense, Spratt (2005: 24) states that: “*the influence of tests can be observed in the curriculum, classroom materials, teaching methods, participants' feelings and attitudes towards the test and finally in the learning process*”.

The concept is derived from the idea that tests and examinations should drive teaching and hence learning. Popham (1987) coins the term “*measurement-driven-instruction*” to describe the mismatch between the content and the format of the test and the current content and format of the curriculum, this is referred to as curriculum alignment by Shepard(1993).

The notion of washback has been described in literature as ‘*test impact*’ (Baker, 1991), ‘*consequential validity*’ (Messick, 1989, 1996), ‘*systemic validity*’ (Frederiksen & Collins, 1989), ‘*measurement-driven instruction*’ (Popham, 1987) or ‘*curricular alignment*’ (Madaus, 1988; Smith, 1991a). Shohammy (1992:513) competes that “*the use of external tests as a device for creating [a positive] impact on the educational process is often referred to as the washback effect or measurement driven instruction*”.

Besides, washback is described in the testing literature as the consequence of the use of tests on school performance(at the micro level) and its effect on the society as a whole(at the macro level) (Andrews, 2004). Oddly enough, when coming to analyze the significance of the word, *washback* is used to refer to the effect of tests on

Chapter One: Background Account of the Selected Literature

teaching /learning at the micro level, whereas effects at the macro level are referred to as *test impact* (Wall, 1997; Hamp-Lyons, 1997; Bachman and Palmer, 2010; Brown and Abeywickrama, 2010).

The washback phenomenon is seen to be restricted to the unexpected but unintentional effects and not to the anticipated effect of tests (Spolsky, 1995). It is described as being any effect, be it positive or negative, intended or unintended, that influences the teaching/learning process as a result of an examination, be it a public examination or a high stakes exam. (Alderson and Wall, 1993; Bachman and Palmer, 1996; Hughes, 2003; Cheng *et al.*, 2004; Cheng, 2005; Bachman and Palmer, 2010; Hung, 2012). Early definitions of the same term may be displayed in the following table:

Authors	Definitions
Buck (1988:17)	<i>“Testing drives not only the curriculum, but also the teaching methods and students’ approaches to learning.”</i>
Alderson & Wall (1993:239)	<i>Concerns has long been voiced about the power of tests to affect what goes on in the classroom, the educational system, and the society as a whole – the so-called ‘washback effect’.</i>
Biggs (1994:12)	<i>“influence of testing on teaching and learning.”</i>
Cohen (1994:41)	<i>“a part of the impact a test may have on learners and teachers, on educational systems in general, and on society at large.”</i>
Bailey (1996:259)	<i>“It is common to claim the existence of washback (the impact of a test on teaching) and to declare that tests can be powerful determiners ... of what happens in classrooms.”</i>
Hugh (2003 : 53)	<i>" an intended or unintended (accidental) direction and function of curriculum change on aspects of teaching and learning by means of a change of public examinations’</i>
Cheng (2005:56)	<i>The influence of the test on the classroom ... this washback effect can be either beneficial or harmful.”</i>

Table1.4. Selected Definitions of Washback

As mentioned above, conceptions of the washback phenomenon are numerous, they vary from plain and straight forward to very complex. Some view washback in terms of teachers and learners in classroom settings, while others insert inferences on educational reforms and even society in general.

In a clear definition about testing in an educational context, Buck(1988) describes the phenomenon as being a natural predisposition for both teachers and

Chapter One: Background Account of the Selected Literature

learners to tailor their classroom activities to the requirements of a test; particularly high stakes exams which have a strong influence on the testers future lives and prospective careers. The focus here is on language learners as test-takers through "*the utilization of external language tests to affect and drive foreign language learning in the school context*" Buck (1988. 513). In the same line of thought, Shohamy (1993:4) summarizes some key definitions underlining the ambiguity of the washback concept:

- Washback effect refers to the impact that tests have on teaching and learning.
- Measurement driven instruction refers to the notion that tests should drive learning.
- Curriculum alignment focuses on the connection between testing and the teaching syllabus.

On his side, Pierce (1992:687), defines it as "*the impact of a test on classroom pedagogy, curriculum development and educational policy.*" Thus, there exist a number of factors influencing the impact of washback in language teaching. In this fashion, Shohamy et.al, (1996:299) states that the degree of impact is influenced by a number of factors, as demonstrated below:

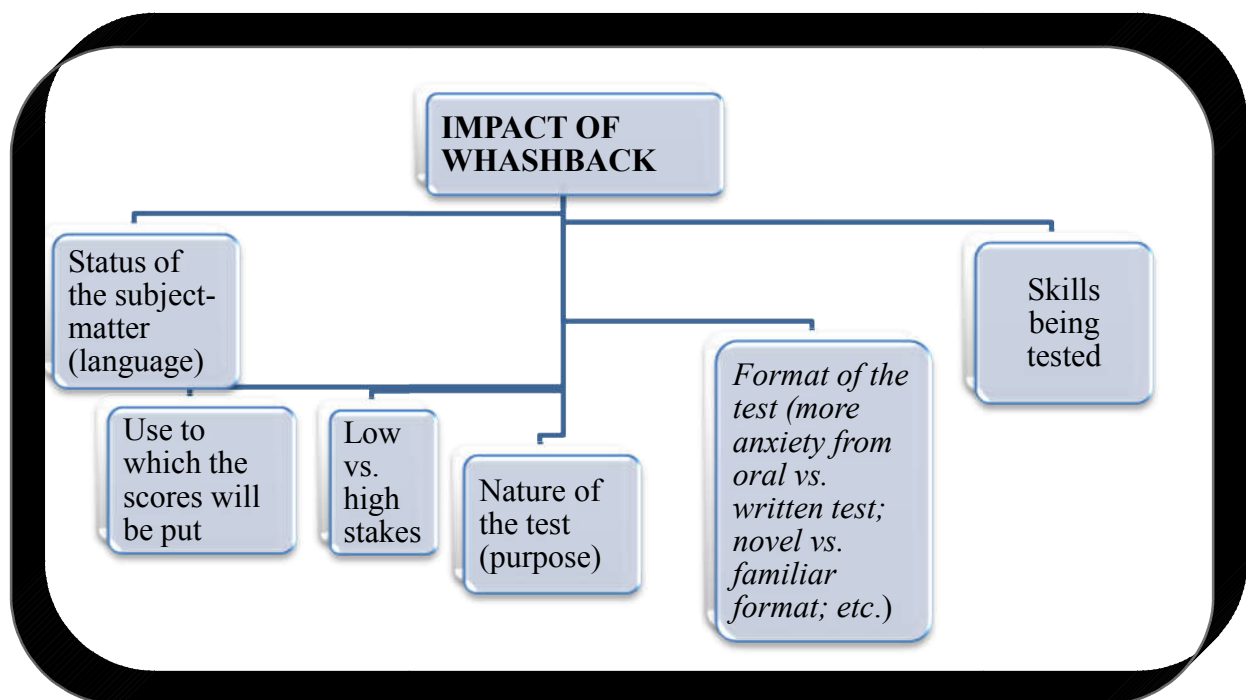


Diagram 1.1. Factors Influencing the Washback (Adopted from Shohamy et al, 1996:299)

Chapter One: Background Account of the Selected Literature

It is a truth generally acknowledged that the influence tests have over curriculum and teaching was seen as a negative one, and the term washback was seen as a destructive effect of tests on teaching/ learning process. The most prominent view is that within teaching, the main objective is practicing test techniques rather than language skills (Vernon, 1956; Wiseman, 1961). In this case, teachers apply practice test after practice test to help learners get familiar with test items rather than develop their competence to manage tests. Wiseman (1961:67) states that: “*An implication of this is that a good test is the one that has no influence on class activities, what Davies calls “an obedient servant that follows its leader, namely, the syllabus and teaching* (Davies, 1985). Herein, teachers tend to direct their classroom activities to meet tests needs, and to render learners slaves to tests.

1.3.2. Washback Dimensions

From the above section, it was clearly found that washback is a highly complex, multidimensional phenomenon (Bailey 1996; Cheng 1997; Watanabe 1996). Hence, it is wiser to check washback dimensions from various aspects to well understand this exceedingly multifaceted phenomenon. Six dimensions are proposed based on a number of studies such as (Bachman and Palmer 2010; Green 2007; Hawkey 2006; Watanabe 2004, Qian 2018). The following table summarises the suggested dimensions:

Dimension	Description
Direction	It denotes if the test may generate positive or negative washback, i.e., if the test is well designed and appropriately used, it will in all probabilities facilitate the teaching and provide constructive feedback, however, it is <i>poorly designed</i> and inappropriately used, it will lead to test failure.
Extent	It denotes the extent of the washback effects, i.e., <i>whether it is the school context or the educational system as a whole. Generally speaking, the more important a test is, the greater its washback effects will be.</i> Qian (2018:56) For instance, a simple classroom

	achievement test may affect only the target class; while a high-stakes test, such as BAC Exam will affect teachers, policy makers, parents and students.
Intensity	It denotes whether washback results strong or weak effects. The intensity of washback is <i>often related to the stakes of a test, i.e., the more important the test, the more intense its washback will be, for instance, a test with more intense washback tends to attract more attention from teachers, learners and researchers.</i>
Intentionality	It indicates if Washback is intended or unintended. Intended washback refers to the effects that teachers can predict or want to encourage, e.g. to motivate students to learn, provide support and improve teaching and learning. Nonetheless, it <i>can bring unintended effects, such as test anxiety among students, negative feedback and long term test-taking technique training.</i>
Length	It denotes the length of the effects, it may last for a short or long term, for instance, <i>motivating students to learn may vanish shortly after the test is over, while developing learning habits to well prepare a test may last a life time.</i>
Specificity	Washback can be general or specific. General washback refers to the effects generated by any test in a general way (broadly speaking). Specific washback occurs only with a certain test, such as the washback effects of oral proficiency. Therefore, in the present study, attention is paid to exploring the specific effects of a test, apart from reviewing the general washback. (Qian, 2018)

Table 1.5. Washback Dimensions

1.3.3. Washback: Positive, Negative, Neither or Both?

Language testing framework has generally been concerned with how tests can influence the way teachers teach and the way learners learn. In view of this, Wiseman (1961:159) states that tests have “*debits*” as well as *credits*. Expressed differently,

Chapter One: Background Account of the Selected Literature

washback phenomenon is bi-directional (Alderson and Wall, 1993), depending on whether this washback has helpful or a harmful effects on the educational process (Hughes,1989). Therefore, washback can be positive or negative, intended or unintended; it has a powerful influence on teaching/learning and may either promote or inhibit the process fluidity. In a more comprehensive manner, Wall and Alderson (1993: 41) clearly state that: *'tests can be powerful determiners, both positively and negatively, of what happens in classrooms'*.

Within a positive scenario, and diving into the existing testing literature at the micro level, if the test promotes effective teaching and thus, creative learning, one can say that the test has a positive washback on the classroom environment. To realize that, Pearson (1988), Shohamy (2001); Brown and Abeywickrama (2010) state that *"for a test to promote beneficial washback, it should be purposive, well-known to teachers and students, as well as reflecting the course objectives upon which the test content is supposedly based"*. In the same line of thought, Messick (1996: 241-242) states that *"for optimal positive washback there should be little if any difference between activities involved in learning the language and activities involved in preparing for the test"*.

On the positive side of the ledger, tests oblige lecturers to inculcate the syllabus with *standards-related content*, thereby creating a pecking order of educational priorities. In this sense, Davies (1985:8) maintains that *'creative and innovative testing . . . can, quite successfully, attract to it-self a syllabus change or a new syllabus which effectively makes it into an achievement test'*. In such a situation, the test no longer needs to be just an obedient servant but rather a leader towards change.

Moreover, some claim that the washback effects force teachers to follow and draw an emphasis on *"drill and kill"* pedagogical techniques; teachers relate that more labs, discussion and critical thinking are occurring in classrooms (Clarke et al., 2002). Admittedly, there should be a kind of balance between the content of the teaching/learning and the content of activities found in the test. Therefore, in high stakes testing, it would be advantageous if the items evaluated overlap with the content of the teaching curriculum (curriculum alignment) (Orafi and Borg, 2009).

Chapter One: Background Account of the Selected Literature

On the other side of the corner, language tests are often judged for their negative influence on teaching. In fact, research has shown *that paid classes* (i.e., classes intended for preparing students for exams) tend to ignore at a large extent subjects and activities that did not contribute directly to examinations (Vernon, 1956, Wiseman, 1961), and this results in passive students practicing test techniques rather than language in-class activities; making the educational experience fine, limited and boring.

At the micro level, teachers confess of losing autonomy when focusing on exam-related-content (specially high stakes exams) and this causes a destructive effect on the classroom settings. For instance, Noble and Smith (1994:6) confirms that '*high stakes testing could affect teachers directly and negatively....teaching test-taking skills and drilling on multiple-choice worksheets is likely to boost the scores but unlikely to promote general understanding*'. Above and beyond, Bailey (1996:269) states that “*...it can be positive or negative, to the extent that it either promotes or impedes the accomplishment of educational goals held by learners*”. Herein, if a test has positive washback, '*there is no difference between teaching the curriculum and teaching to the test*'. Weigle & Jensen (1997: 205). However, if it has a negative washback, it may lead to the abandonment of instructional goals for test preparation, what is generally known as the 'Teach-to-the-Test' Approach (see Section 1.3.4. for more information).

Using Pearson's (1988:101) words: '*a test's washback effect will be negative if it fails to reflect the learning principles and course objectives to which the test supposedly relates, and it will be positive if the effects are beneficial and “encourage the whole range of desired changes*'. Consequently, tests will become inappropriate practices; they will have negative effects on teaching/learning when learners' test scores rise without a genuine development in learning causing “*test scores pollution*” Haladyna (2001:20). Furthermore, this promotes traditional ways of delivering knowledge, creating '*dull teaching*'. Goruch(1999:25) describes it as:

- “a) teacher centered;*
- b) teacher-to-whole-class oriented;*
- c) focused on the learning of discrete facts;*

d) Product oriented in that students are expected to repeat facts through recitation and written tests”.

Therefore, following such a scenario evokes a traditional teaching approach in which students will learn through rote learning and memorization rather than reflection and creative learning, forcing teachers to focus on the quantity rather than the quality, and on scores rather than performance (Black and Wiliam, 2006). At this very crucial level, it seems to be wiser to refer to teach-to-the test approach as one of the crucial results of negative washback.

1.3.4. Teach-To-The-Test Approach

Several attempts have been carried out speaking about the effect of the testing on the teaching process. Therefore, in a more atomistic analysis of the current situation, one may witness a mismatch between the settled course goals and objectives and a clear emphasis on assessment within the course of instruction. Thus, emphasizing on tested subjects leads teachers to focus more on exam-related-content and decreases emphasis on non tested subjects and narrows curriculum accordingly. In this sense, Kaufman(2015 :3) states that *“All of this time spent preparing for standardized tests and actually taking the tests, means that teachers have to focus on subjects that are tested and teach test-taking skills”.*

Within the same line of thought, teachers applying this, will tailor their classroom practices to meet exam necessities and get better students' scores, *“this impairs educational quality by distorting the curriculum, and trivializes some important aspects of language learning, i.e. narrowing the curriculum ”* (Shohamy, 2001; Cheng and Curtis, 2004; Saif, 2006: 245).

This phenomenon seems to impact teachers negatively in devaluing their professional status for assessment. Teachers working out with their students a whole year long on a daily basis are in better position to know their students' abilities and achievements and thus design the test accordingly as it is provided by

Chapter One: Background Account of the Selected Literature

Kaufman(2015:5): *“Given this wealth of data collected over time, the teacher is better positioned to evaluate student achievement than a one-off high-stakes test.”*

Additionally, high-stakes testing may not reveal students’ competence. One reason is that students may experience a test anxiety and perform poorly on standardized tests , as it is put by Kaufman (2015:7) *“ Either way, high-stakes testing around the world leads to intense levels of stress and anxiety for students given that all of their hard work over the years comes down to one test”*.

Besides, in a recent press conference about ‘*The Danger of Standardized Tests*’, Chomsky (2015) roughly argued about the danger of standardized testing on students’ potentials and learning ambitions he states that the *“assessment itself is completely artificial”* and that:

The people sitting in the offices, the bureaucrats designing this, they’re not evil people, but they’re working within a system of ideology and doctrines that turns what they’re doing into something extremely harmful.

Given the variety of reasons that leads to anxiety, preparing students for a high stakes examination is, in fact, a great responsibility since it has a great determinant of students’ future life and prospective career. In this fashion, Valerie states that:

It was very clear for everybody that unless you do very well with this one examination, that some of these dreams that you may have for the future will become very difficult to fulfill.

Valerie (2014:24)

High stakes tests are designed without giving consideration on how students will feel as human beings. They are treated like *industrial products* that must be adjusted to control quality before they are out for public use. It demeans students’ potentials and ambitions to grow intellectually in this ever changing time of globalization (Chomsky,2015).

In a more intricate manner, researchers agree on the fact that testing appears to be anxiety-provoking, and teachers need to consider the psychological side of learners when being tested. Odo states that:

Researchers are becoming progressively more aware of the negative social impact large-scale high-stakes tests can have on the lives of learners – particularly those who are most vulnerable – when the results of these tests are used to make decisions that unfairly limit the life choices of these learners. Odo (2012:2)

Thus, since results of these high stakes tests will control students' future lives and prospective career, teachers must follow corrective methods in an attempt to fix the breaks resulting from their testing approach. Therefore, the subsequent section will introduce the different hypotheses proposed by a number of researchers, this will offer a clear picture that will help us better understand the phenomenon to better investigate it.

1.3.5. The Washback Hypothesis

A genuine importance is connected to the mechanism of washback functions. However, it is assumed that little proposals have been made regarding *the type and organization of the factors interacting with the tests, to bring about beneficial washback*, Bailey (1999:25).

In his unpublished paper (cited in Bailey, 1996), Hughes (1993) suggests an authentic model of washback which can be built grouping their versatile distinctions. In this sense, Alderson and Wall (1993, cited in Benmostefa, 2014) put forward the following hypotheses:

1-A test will influence teaching.

2-A test will influence learning.

3-A test will influence what teachers teach.

4-A test will influence how teachers teach.

5-A test will influence what learners learn.

6-A test will influence how learners learn.

7-A test will influence the rate and sequence of teaching.

8-A test will influence the rate and sequence of learning.

Chapter One: Background Account of the Selected Literature

9-A test will influence the degree and depth of teaching.

10-A test will influence the degree and depth of learning.

11-A test will influence attitudes to the content, method, etc. of teaching and learning.

12-Tests that have important consequences will have washback.

13-Tests that do not have important consequences will have no washback.

14-Tests will have washback on all learners and teachers.

15-Tests will have washback effects for some learners and some teachers, but not for others

Alderson and Wall (1993: 120-121)

In a more detailed manner, a number of researchers Hughes (1993), Bailey(1994) designed comprehensive models for better understanding of washback phenomenon they are put as follows:

- Hughes Model 1993

As mentioned in Bailey (1999), Hughes (1993) argued that the fifteens hypotheses were quite general, so it would be preferable to precise a better exam washback presentation “*In order to clarify our thinking on backwash, it is helpful, I believe, to distinguish between participants, process and product in teaching and learning, recognizing that all three may be affected by the nature of a test*” (cited in Bailey, 1999: 9).He puts forward a tracheotomy as displayed in the following diagram:

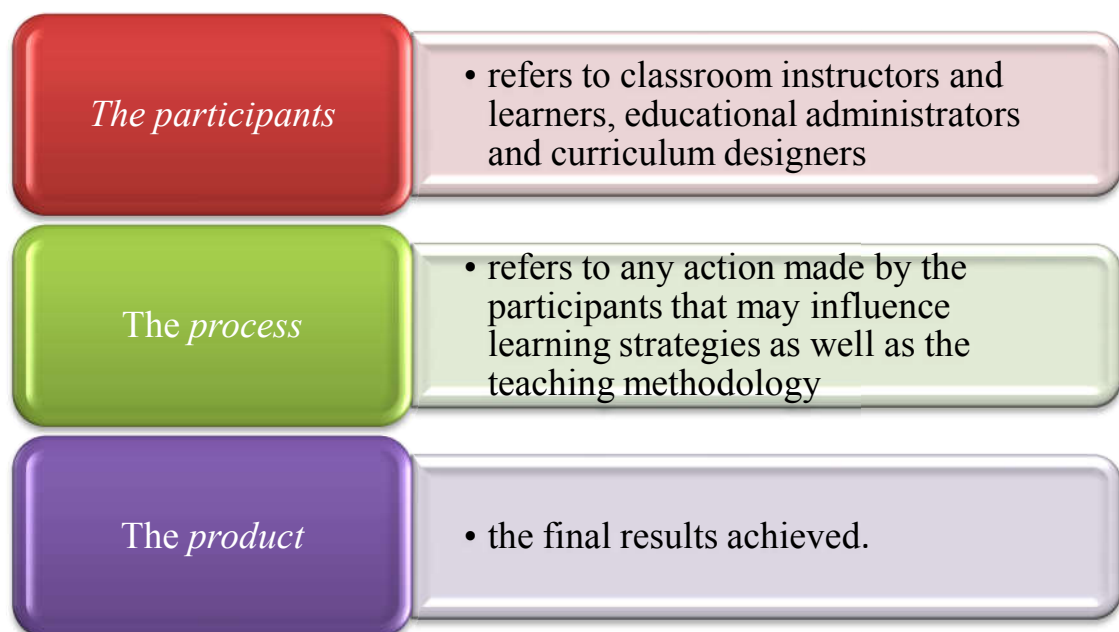


Diagram 1.2. Bailey’s Trichotomy (1999:9)

Chapter One: Background Account of the Selected Literature

The result of this complexity may bring into question the evaluation of the effectiveness of the educational system. Hence, Hughes (1993:2) explains trichotomy washback model as follows:

The trichotomy [...] allows us to construct a basic model of backwash. The nature of a test may first affect the perceptions and attitudes of the participants towards their teaching and learning tasks. These perceptions and attitudes in turn may affect what the participants do in carrying out their work (process), including practicing the kind of items that are to be found in the test, which will affect the learning outcomes, the product of the work.

In an attempt to establish the connection between Alderson and wall’s Hypothesis as well as Hughes tracheotomy, the following table shows this connection:

Hughes Trichotomy	Washback Hypothesis
Participants Process	A test will influence what teachers teach. -A test will influence how teachers teach. -A test will influence what learners learn. -A test will influence how learners learn.
Product	-A test will influence teaching. -A test will influence learning. -A test will influence the rate and sequence of teaching. -A test will influence the rate and sequence of learning. -A test will influence the degree and depth of teaching. -A test will influence the degree and depth of learning. -A test will influence attitudes to the content, method, etc. of teaching and learning. -Tests that have important consequences will have washback. -Tests that do not have important consequences will have

	<p>no washback.</p> <p>-Tests will have washback on all learners and teachers.</p> <p>-Tests will have washback effects for some learners and some teachers, but not for others.</p>
--	--

Table1.6.Connection between Hughes’ Trichotomy and Alderson and Wall Hypothesis (1993).

- Bailey’s Model:

Combining Alderson and Wall’s hypothesis with the trithology of Hughes’ (1993), Bailey (1996) proposed a mid ground model. She competes that a test not only affects the product through the participants and their actions(processes), but that participants can in turn provide feedback and then, this, can have an impact on the new test. This suggestion can be summarized as follows:

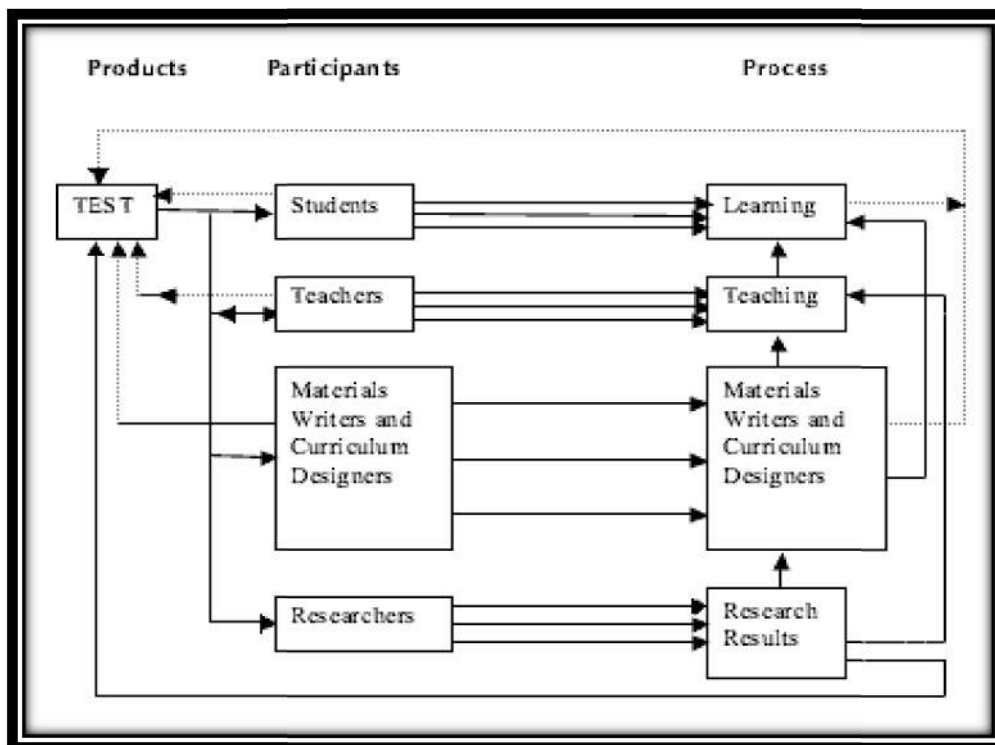


Figure1.1. Models of Washback (Bailey,1994:264)

1.3.6. Previous Related Studies on Washback

Factors influencing language teaching and language assessment are in vogue within recent research worldwide. The washback phenomenon has been under the inspection of many educationalists and didactitions for several years, each language researcher approaches it differently. Early studies within the same area of research may be displayed in the following table:

Researcher	Study	Results achieved
Shohamy's (1993)	Studied washback effect on three language tests (the EFL oral test, ASL test and L1 reading test).	the methodology teachers adhered became more "test-like" as the exam approached
Wall and Alderson's (1993)	Their study examined the effects of introducing a new English examination that was intended to encourage a more communicative approach to teaching,	It revealed that the examination "has had virtually no impact on the way that teachers teach English"
Watanabe's (1996)	Based on classroom observation of two teachers,	He found that they were influenced by the test in terms of how they teach, but the degree of this influence contrasted from one teacher to another. Teachers changed their teaching methods and techniques toward the exam.
Watanabe (2000),	investigated washback effects of a pre-college English section examination on instruction in Japan through classroom observations and teacher interviews	reported that teachers in his study "claimed that they deliberately avoided referring to test taking techniques, since they believed that actual English skills would lead to students' passing the exam" (45).
Stecher et al. (2004),	He investigated the influence of WASL tests on methods teachers used in teaching	He found that teachers changed their methods to reflect test requirements.

	writing,	
Cheng, (2005)	Based on Questionnaires, classroom observations, interviews, document analysis with policy makers and teachers,	<ul style="list-style-type: none"> - The revision of a test will lead to efficient washback on teaching materials; - Teaching content will be affected, but teachers' attitudes and behaviours will not change much. <p>A change in the test only can hardly fulfil the intended goal of improving teaching/ learning</p>
Wall (2005),	Administering questionnaires and interviews to teachers in addition to classroom observation to check the influence of washback on teachers focusing on the factors involved.	<ul style="list-style-type: none"> -Tests have washback on what teachers teach, but not how they teach, and there is both positive and negative washback; -Many factors contribute to or inhibit washback, implying that the nature of curricular innovation is much more complex
Amengual-Pizarro (2009)	She examined the influence of the ET test included in the SUEE on various aspects of teaching.	"The results of this study also appear to indicate that the ET affects the methodology teachers employ in actual class teaching adapting it to the purpose of the test" (2009:594).

Table1.7. Early Studies on Washback.

1.3.7. A Critique of the Washback Studies

Washback study is generally described as comprehensive and systematic operation within micro context (the school setting) or in the macro context (the educational system or the society). It greatly involves various actors such as test takers, teachers, parents, policy makers, society pressure... Hence, it seems quite impractical to create a unique absolute approach which covers all aspects of washback Qian (2018).

Conclusions from the previous studies (see Table 1.5.) denote that washback researchers were basing their attention mostly on '*large-scale, multi-method, multi-phase and longitudinal*' Qian (2018:31). This is described by the fact that the investigated tests are most of the time high-stakes tests which are affected by a wide

range of variables. However, researchers' main concern was the teacher and the learner as the unique heroes within high stakes testing, neglecting, at a certain extent, its infinite stakeholders.

Besides, studies on washback according to the previous studies are often based on longitudinal research since they employ various methods within the teaching/learning process. However, washback effects require much time to emerge, for instance, Qi (2004) concludes his research by stating that: *'high stakes tests can hardly realize their intended washback because of the overwhelming functions they are expected to perform'*.

1.4. ORAL PROFICIENCY IN EFL CLASSROOM

Diving into the history of the discourse in education, one may notice an interesting focus on learners' achievements. In this globalised world, the need for a competent, highly educated, well informed learner has been the concern of educationalists for a long period of time. Foreign language learners generally study foreign languages wishing to become fluent and get a native like pronunciation, including mastering different language structures and vocabulary. However, language researchers still doubt about what constitute the oral proficiency.

Characteristics of proficient language speakers are called "good" "fluent" and "competent"; however, it is not always clear what oral proficiency clearly denotes. The term has been used differently by different researchers each according to his general framework of research. Over four decades, oral proficiency has been promoted to be part of the dominant competences in language education. Communicative competence has been at the center of the English language education syllabi worldwide. Oral communicative proficiency is the essence of language proficiency.

1.4.1. The Notion of *Oral Proficiency*

The concept of '*oral proficiency*' has been central to foreign language learning and research for the past several decades. Numerous definitions of oral proficiency have been put forward; they differ among researchers making it difficult to set down for a specific definition, (McNamara, 1996). Multiple characteristics have to be taken

Chapter One: Background Account of the Selected Literature

into consideration, among which: grammar, vocabulary, fluency, pronunciation, comprehensibility are to be included.

It has been argued for many years now that language proficiency denotes grammatical proficiency *per se* (Leclercq: 2014). However, in 1972, Hymes argued that the communicative language use is equally important as the grammatical use. Lado (1967;34) on his side expended the idea of proficiency and further adds the four language variables, pronunciation, grammatical structure, lexicon and cultural significance, in addition to the four language skills , (cited in Young & He, 1998). Canale and Swain (1980) further defines language proficiency and adds linguistic and pragmatic competence, discourse and strategic competence as key components of oral proficiency (as cited in Young & He,1998).

Skehan (1989) was the pioneer to define proficiency with three core components, Complexity, Accuracy, and Fluency as known as CAF. Complexity is how varied and elaborate the speakers' language is. Accuracy naturally means correct language use. Eventually, fluency, is defined as how closely speech resembles the native speakers'. Comprehensively put, Ellis and Barkhuizen (2005: 139) '*Fluency has been defined as the production of language in real time without undue pausing or hesitation*'), some view it as how effortlessly and 'fast' language is produced.

According to ACTEFL proficiency guidelines (2012), oral proficiency serves different language functions among which the following:

- Proficient language speaker has the ability to adapt language to specific audience, negotiate meanings, and deal with vocabulary nuance.
- Discuss different topics, support opinions, hypothesis
- Narrate and describe in different language tenses.
- Bring about simple conversations by asking and responding to simple questions.

1.4.2. Testing the Oral Proficiency

With the global trend towards cross cultural education, linguistics diversity and internationalization, language assessment has become more persistent and more powerful decision-making tools (Shohamy, 2007). It is generally believed that assessment is an integral part in didactics. It is addressed in different formats on different language variables. However great the emphasis on assessment is, little is done concerning language testing purposes and some parts still remain an enigma within the literature. It has been suggested that teachers around the world have little or no competence when assessment and grading are concerned (Lundahl: 2011). According to a research carried out in a Swedish department of foreign language teaching, Swedish teachers claim to have no previous background on assessment (Lundgren & Nihlfors, 2005). Thus, teachers feel the need for more training especially when it comes to assessment and grading.

Speaking is regarded as the most complex skill to assess. It involves a mixture of skills that may have no correlation with each other, and which do not lend themselves well to objective testing (Kitao & Kitao:1996). In this vein, Lado exhaustively posits:

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...there is a clear lack of understanding of what constitutes speaking ability or oral production.

Lado (1961: 239)

Thus, speaking skills is regarded as one of the most complex and controversial aspects within languages teaching/testing (O'Sullivan, 2006), speaking is, thus, considered as the hardest skill to teach, practise and test. What makes it more complicated is that scores are not offering reliable information on learners' real competence, since they offer a limited picture on their '*speaking ability*' rather

Chapter One: Background Account of the Selected Literature

'speaking ability in x, y or z context', Glen (2003:19). In other words, test purpose should take into account other dimensions and factors.

As for assessment and scoring of speaking skill, recording learners' seems to be the most adequate assessment method, and the scoring phase will be done through listening to the recorded tape. The assessment strategy includes grammar, pronunciation, fluency, content, organization, and vocabulary. (Kitao & Kitao:1996).

It has been argued that accomplishing one aspect of oral interaction may hinder the ability to attend the other aspects (Krashen,1992). In assessing speaking, one should pay attention to the limited time given for planning, the vocabulary used, and grammar of the spoken language. Thus, these criteria should be taken into consideration when assessing. In this vein, Shumin, (2002:204) states that:

In order to provide effective guidance in developing competent speakers of English, it is necessary to examine the factors affecting adult learners' oral communication, components underlying speaking proficiency, and specific skills and strategies used in communication.

Hence, assessing oral proficiency and interaction in a collaborative action is a daunting task. To reach effective assessment, teachers and test takers need to share global knowledge. Language instruction aims to develop proficiency in four known areas: written language, reading proficiency, listening ability and oral language production (National Standards in Foreign Language Education Project, 1999). The first three areas are willingly measurable through traditional paper-based assessment, such as written exams. The assessment of the oral production is a real challenge facing today's educators. Thus, many attempts have been put to find out the magical recipe for testing oral proficiency adequately. For instance, a number of criteria are suggested to measure learners' fluency. These criteria are displayed below:

Criteria	Explanation
-----------------	--------------------

Speech Rate	Total number of syllables divided by total time taken to execute the oral task in hand
Mean length of run	Average length of syllables produced in utterances between short pauses
Phonation/time ratio	Time spent speaking divided by the total time taken to execute the oral task;
Articulation rate	Total number of syllables divided by the time to produce them

Table1.8. Assessing Fluency(adopted from Towell,2006)

1.4.3. Problems in Testing the Oral Proficiency

Testing the oral proficiency seems to be a complex task that requires careful attention from the part of teachers to attain the necessary objectives and goals. The reasons behind such a challenge are put by Sujana (2016) as follows:

- The nature of speaking skill is quite complicated;
- teachers find it difficult to select what to test (the criteria in testing speaking ability);
- the influence of other factors such as listening ability, pronunciation ability and reasoning ability;
- the difficulty in getting students to speak.

Teachers often find it difficult to assess the oral proficiency due to the complicated process of learning speaking in particular. Problems encountered can be put in the subsequent table:

Researcher	Problems encountered
(Ur 1995: 121)	<ul style="list-style-type: none">• inhibition – fear of making mistakes, losing face, criticism; shyness;• nothing to say – learners have problems with finding motives to speak• formulating opinions or relevant comments;• low or uneven participation – often caused by the tendency of some learners to dominate in the group

Chapter One: Background Account of the Selected Literature

	<ul style="list-style-type: none">• mother-tongue use – particularly common in less disciplined or less motivated classes, learners find it easier or more natural to express themselves in their native language.
Morrow (1982)	<ul style="list-style-type: none">• Oral testing is very time-consuming,• It is difficult to get students to say anything interesting;• Teachers expect learners to entertain them with brilliant conversation but they should encourage them at least to use the language for a variety of purposes (describing, narrating, apologizing, etc.); being able to take part in spontaneous conversation, responding appropriately, making relevant contribution; and having the chance to show that he can perform linguistically in a variety of situations,• oral proficiency tests are hard to score objectively, teachers always find it challenging to choose the appropriate criteria

Table 1.9. Problems of Oral Proficiency Assessment

In order to eliminate those problems, Morrow (1982) further suggests these solutions that might be helpful for teachers:

- The tasks designed should be as close as the real world;
- Setting group work. The group work can at least solve the problems related to the time consuming issue and give a chance to students to use the language spontaneously, involving a variety of functions;
- Setting clear criteria for the examiner.

Therefore, current research (Vogt and Tsagari, 2014; Tsagari and Vogt, 2017) has shown that English language teachers often lack sufficient understanding of the nature of assessment in general and are not familiar with the relevant assessment techniques and methods, in particular to renew, change and develop. Thus, it is

important to enhance teachers' level of Language Assessment Literacy (LAL) (Inbar-Lourie, 2008), that is, their "*ability to design, develop and critically evaluate tests and other assessment procedures, as well as the ability to monitor, evaluate, grade and score assessments on the basis of theoretical knowledge*" (Vogt & Tsagari (2014: 377).

1.5. INVESTIGATING THE AMBIGUITY OF WASHBACK PHENOMENON ON LEARNERS' ORAL PROFICIENCY

Looking deeply at the content and approaches used in assessment practices in higher education, one may diagnose a profound gap between delivering knowledge and assessing it. Major focus is placed on students' grading and scoring, neglecting at a certain extent on the processes of teaching, learning and assessing how students will learn after feedback. In other words, assessment is not sufficiently well designed, equipped and tracked.

A truth that is diagnosed throughout this background account is that despite the fact that washback has long been a concept under discussion (Cronbach 1963; Fredericksen 1984; Latham 1877), experimental studies on the washback effects of language testing is a newly area *per se*. Within the field of language testing, researchers' main interest was to investigate issues and problems related to tests and how to increase its reliability and validity. However, if one considers washback and its ambiguity, it goes well beyond the test itself. Teachers, researchers and policy makers should consider '*the plethora of variables, including school curriculum, behaviours of teachers and learners inside and outside the classroom, their perceptions of the test, how test scores are used, and so forth*' Cheng (2008: xiv).

Contextualising this view, and considering the ambiguity of washback on learners' oral proficiency, it is to be noted that examinations, especially high-stakes tests, have powerful washback effects on teaching and learning within different educational context (Cheng, 1998; 2018). In fact, assessment can leverage educational systems, and the ambiguity of washback will remain questioned, calling for research and empirical investigations to limit its negative effects.

Chapter One: Background Account of the Selected Literature

Language tests seem to have a more direct washback effect on teaching process, despite the increasing literature on washback, experimental studies are rather little and this causes its ambiguity. Thus, in order to develop a comprehensive understanding of the phenomenon in general and on oral proficiency in particular, one should investigate the matter from different aspects and establish possible relationships between testing, teaching, and learning in a classroom setting. This is, in fact the concern of the subsequent chapter.

1.6. CONCLUSION

This chapter aimed at bringing the theoretical part of this research, it tried to highlight the relevant literature needed to explain the research aim and objectives put at the beginning, shedding light on key concepts related to testing and foreign oral proficiency. Attention was also put on the relationship between testing, the washback phenomenon and oral language proficiency. This account was to prepare the ground for the second chapter.

Chapter One: Background Account of the Selected Literature

Chapter Two: An Outlook into Research Design

2.1. INTRODUCTION.....	45
2.2. LANGUAGE POLICY IN ALGERIA: REFORMS AND TESTING...	45
2.2.1. English Language Testing at the Tertiary Level.....	47
2.3. RESEARCH PHILOSOPHY AND PARADIGM.....	50
2.3.1. Research Roadmap, Aims and Research Questions.....	54
2.3.2. Research Timeline and Phases.....	55
2.4. SAMPLING ADMINISTRATION.....	55
2.4.1. Students' Profile	56
2.4.2. Teachers' Profile.....	56
2.5. RESEARCH INSTRUMENTS.....	56
2.5.1. Cycle one	57
2.5.1.1. Needs Assessment Questionnaire: Pre-Teaching Proficiency Test	57
2.5.1.1.1. One-Sample Kolmogorov-Smirnov Test.....	63
2.5.1.2. Washback Questionnaire.....	64
2.5.1.3. Learners' Oral proficiency Test.....	68
2.5.1.4. Oral Proficiency Rating Scale.....	70
2.5.1.5. Teachers' Interview.....	71
2.5.1.6. Classroom Observation Phase I.....	74
2.5.2. Circle Two.....	75
2.5.2.1. Washback Questionnaire II.....	76
2.5.2.2. Classroom Observation Phase II.....	78
2.5.2.3. Oral Proficiency Test.....	79
2.5.2.4. Oral Proficiency Rating Scale.....	80
2.5.2.5. Post-Test Questionnaire.....	81
2.6. Conclusion.....	84

2.1. INTRODUCTION

Based on the previous theoretical chapter settled which aimed at building up the stones for the present practical chapter, this part appears to be descriptive in nature. The query starts with a relevant account of the Algerian testing policy and the status of LMD system at the University of Tlemcen. Furthermore, it includes the research design, data collection method, the selection of the participants; it also intends to present the different research tools used in the current study. Furthermore, this chapter intends to reflect upon action research in foreign language education as a method which may enhance our understanding of how to put into practice effective ways to improve testing learners' oral proficiency.

2.2. LANGUAGE POLICY IN ALGERIA: REFORMS AND TESTING

Algerian educational system has gone through a series of changes and reforms. These winds of reforms blow over the Algerian educational system to modernize and upgrade the existing practices, adopting a newly methodology that suits the 21st century skills needed.

As schools of thought have come and gone, the Algerian educational framework has witnessed a slow but deliberate shift and progress regarding its curriculum development and teaching methodologies. A number of reforms have been introduced to prepare students to take part in the economic growth of the country in the light of intensive modern research and on-going globalization process.

Language teaching/ testing has been largely influenced by these theoretical principles, and there are a set of shortages and negative aspects regarding its implementation, for instance, teachers seem to be not formed to cope well with these alterations. In this vein, Miliani (2010:71) conceives that: *'This new development at school level has generated uneasiness of teachers who are supposed to teach through it but know nearly nothing about it'*. This implies either that the theoretical suggestions are not applied in the right way or they are inappropriately adapted to the Algerian context.

The reservation of testing policies in Algeria seems to be among the most fundamental issues occupying language education research. Therefore, it is difficult to

speak about educational reforms and revolutions without speaking about reforms in the testing arena.

In fact, state policymakers, school leaders, educational authorities, and residents of various kinds are compelled to settle on choices about how to plan and react to new testing frameworks, what motives ought to go with those tests, and what instructors need to know to ensure that testing helps learners develop and grow instead of keeping them getting to be slaves to tests Kaufman (2015).

Teachers generally use tests to evaluate students' strengths and weaknesses and state testing seems to offer a real problematic theme. Notwithstanding the importance of High-stakes testing on teaching process, it has become pervasive in the current educational culture, and stakeholders are all impacted by the pressure to succeed on standardized tests Kaufman (2015). In view of this, he states that:

While standardized tests may seem to offer an excellent way to accurately assess students, It is believed that high-stakes testing leads to numerous negative effects that impact not only the students who take the tests, but also parents, teachers, and schools.

Kaufman (2015:3)

- *High Stakes Testing*

Speaking about high stakes testing, Algeria like many other countries has achievement tests that play an important role in students' prospective lives. The BEM (Brevet d'Enseignement Moyen) and the baccalaureate, end of year examinations in the Algerian educational system Pupils attending the fourth year at the middle school will get access to secondary education and those attending 3 year secondary school will shift from secondary education to tertiary education. (Benmostefa,2014)

The Baccalaureate (BAC) exam, as a high stakes exam, allows the movement from secondary education to tertiary level. It opens the doors for learners to embark in a different learning style and prepare themselves for the creation of their future. It is considered as an important achievement test that changes the lives of learners and seems to be always under the pressure of reforms, this is stated by Benmoussat (2018:02): *“the Baccalaureate examination has always been a lever for change to*

Chapter Two: An Outlook into Research Design and Methodology

initiate educational reforms in an effort to establish academic standards and a more or less reliable source of information on educational outcomes to external stakeholders''.

In this vein, teachers, parents, stakeholders, and policy makers are always under this pressure as it became 'the nightmare' of all learners and parents alike. Oddly enough, and due to the importance of this exam, a new philosophy of testing became fashionable. The teach-to-the-test tendency turns teaching towards scoring-based teaching and not competence-based teaching, and teachers consciously or unconsciously centre their teaching efforts on preparing learners for tests for scoring purposes rather than for competence and proficiency purposes.

This choosy selective focus on exam related content may impact teachers' methodology being on the pressure to teach *Exam English*. Debatably, emphasis should be put on academic success and concentration on the pupil's development of his or her potential rather than on training them for a particular exam (Benmoussat, 2018).

2.2.1. English Language Testing at the Tertiary Level

If one takes an X-ray on the educational dynamics in Algeria, one would find diversity and conflict between what is planned, reality and context. Our educational system seems to be based on quantity-based teaching rather than quality-based teaching. (Miliani, 2010)

Language testing in higher education is extremely widespread; it entails different angles according to the intended teaching objectives. Taking the university as a concrete example, one may think straightforwardly of the word *obscurity*. Despite the fact that teachers normally are provided with the main course content, testing appears to be at a crossroad position. Teachers do not work in harmony with each other, testing takes place differently though within the same line of teaching, scoring has no established rules.

Faced with considerable pressure, it was of great necessity and urgent need for the Algerian educational system to cope with the global expectations and adopt the

Chapter Two: An Outlook into Research Design and Methodology

new directions and global trends of higher education. The LMD system was one of the most important adoptions Algeria went through during the last decade. In this fashion, (Bouhadiba,2013) claims that “*The LMD reform was launched as a pilot scheme in the Algerian universities during the academic year 2003-2004. Three years after the implementation of the first cycle (Licence degree) and at the time of the implementation of the second cycle (Master degree)*”.

Like many countries, Algeria applied the LMD system to meet her learners’ 21st required skills within this globalised age. In fact, Algeria adopted the LMD system as a *prêt-à-porter* European system that may answer its educational current needs. Among the most important reasons behind this adoption the following:

- Improving the quality of instruction in higher education,
- Compatibility of the training with the global system and European standards in particular,
- Diversifying training paths and linking them to the economic and social needs,
- Seeking to ensure employment,
- Modernizing management and pedagogy.

Mezian and Mahi (2010:271)

In fact, testing is challenging at this very particular context due to many reasons. Among these restrictions the following:

Obstacle	Description
<i>Teaching/testing Materials</i>	Observing closely the existing teaching materials, they seem to be traditional, based on the board and the chalk. It is noticed that within testing, there is a total absence of audio-visual aids, laboratories and technological tools which may facilitate the testing process and motivate learners for better scores.
<i>Overcrowded Classrooms</i>	Reflecting on an Algerian context, a classroom is generally made up of 30 to 45 learners, this overcrowded setting may hinder the process of teaching, hence testing. What is more, testing using technological-based tools is

Chapter Two: An Outlook into Research Design and Methodology

	quite impossible within such a context. In fact, within the LMD system, it is supposed that the number of learners do not exceed 20 per group so that they receive the needed care and attention.
<i>Teachers as Test Designers</i>	Teachers rely on their personal learning/teaching experience in designing their courses. Most of them teach/test the way they have been taught with limited motivation, innovation and change. In fact, teachers should be trained how to test and what to test as there is no teaching without testing.

Table 2.1. LMD System Restrictions

Besides, a number of shortages within LMD principles, performance and results are detected. Such shortages are put by Djekoun (2006, qtd in Mezian and Mahi, 2010:271) as follows:

- A large number of students with poor attendance.
- A great failure and dropout rate along with cost effectiveness.
- Weak dynamics when it comes to program renewal.
- Weak relations between the university and its social and economic environments.
- Strong centralization as a means of managing university life.

In a more contextualised fashion, teaching speaking skills in the English department at Tlemcen University is designed within the early three years of instruction. Lessons are generally based on developing learners' general oral proficiency and teachers rely on some textbooks harmonized with handouts and communicative activities (oral presentations, role-plays, interviews, dialogues, etc.). Besides, it is worth pointing out that teachers are free to select what to teach based on diagnosing learners needs and preferences.

Regarding oral proficiency assessment, continuous control throughout the year and a final exam along each semester is programmed for the three years of instruction. Students in such a system can compensate for the courses belonging to the same unit and between different units. Oral tests aim at assessing oral proficiency and listening skills too.

The main concern of this research work is first to examine the testing process of oral courses at the department of English at Tlemcen University, it also tries to see the washback effects on an oral course, and proposes adequate technology-based testing to cope with the needs of the 21st century education.

2.3. RESEARCH PHILOSOPHY AND PARADIGM

It is assumed that attention has grown gradually in research in second and foreign language learning and teaching. This attention is reflected upon the increased professional activity imitated in the growing number of books, journals and conferences devoted to issues of research. The backbone of any research investigation is the selection of appropriate research philosophy and paradigm. In this sense, Saunders et al., (2009: 108) state that '*The research philosophy you adopt contains important assumptions about the way in which you view the world. These assumptions will underpin your research strategy and the methods you choose as part of that strategy*'.

Thus, selecting the appropriate paradigm is an essential step towards successful research. In general, this research is based on an action research fused with several methodologies in order to build a comprehensive answer to a problematic settled. This research work is, then, an attempt to investigate the ambiguity of test washback on learners' general oral proficiency. It relies, therefore, mostly; on qualitative/quantitative methods while integrating some correlational, statistical and survey techniques as well.

The research paradigm followed is an Action Research (AR), it involves '*taking a self-reflective, critical, and systematic approach to exploring your own teaching contexts*' Burns (2010:2). The central idea is to conduct a plan for action for problem-solving situations and problematic. Burns, (1999: 5) comprehensively define it as:

Chapter Two: An Outlook into Research Design and Methodology

A reflective, systematic and critical approach to enquiry by participants who are at the same time members of the research community. The aim is to identify problematic situations or issues considered by the participants to be worthy of investigation in order to bring about critically informed changes in practice.

Thus, for teachers who are reflective, AR is the best way to boost their teaching process, lead to positive change in the classroom, and show teachers' personal approaches in teaching. According to Kemmis and McTaggart (1986), four main phases are followed in developing an AR, the first cycle involves *planning, acting, observing* and *reflecting*, and is a continuing spiral of cycles until the researcher will achieve the satisfactory outcome, the following figure illustrates these cycles:

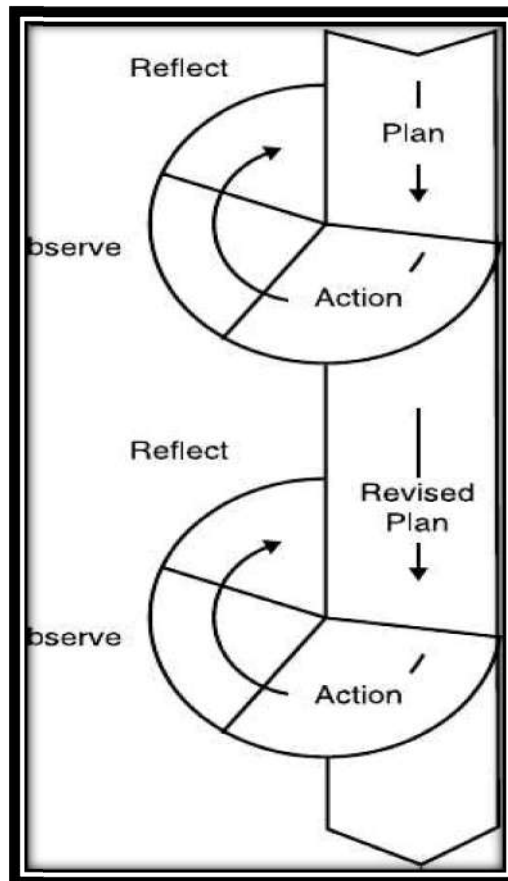


Diagram 2.1. Cyclical AR Model (Kemmis and McTaggart, 1986).

Chapter Two: An Outlook into Research Design and Methodology

Hence, the present research work is intended to investigate the effects tests have on both the teaching of speaking and the overall learners' speaking proficiency. This study is carried out to rise teachers' awareness of the washback phenomenon, and will accompany learners throughout the term with facilitating strategies that will help them develop their speaking proficiency. The following adopted diagram indicates the set of steps a researcher should follow in action research paradigm:

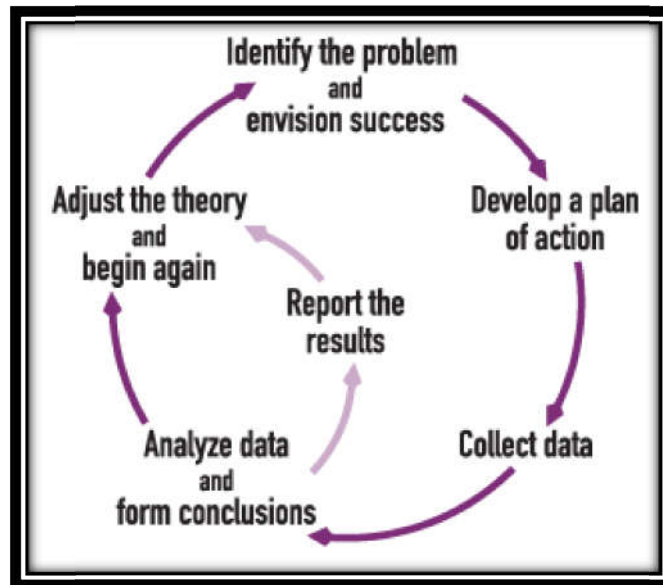


Diagram 2.2. The Process of Action Research (Kemmis and McTaggart, 1986).

To reach this cycle, the researcher will follow the following steps of conducting an action research:

Chapter Two: An Outlook into Research Design and Methodology

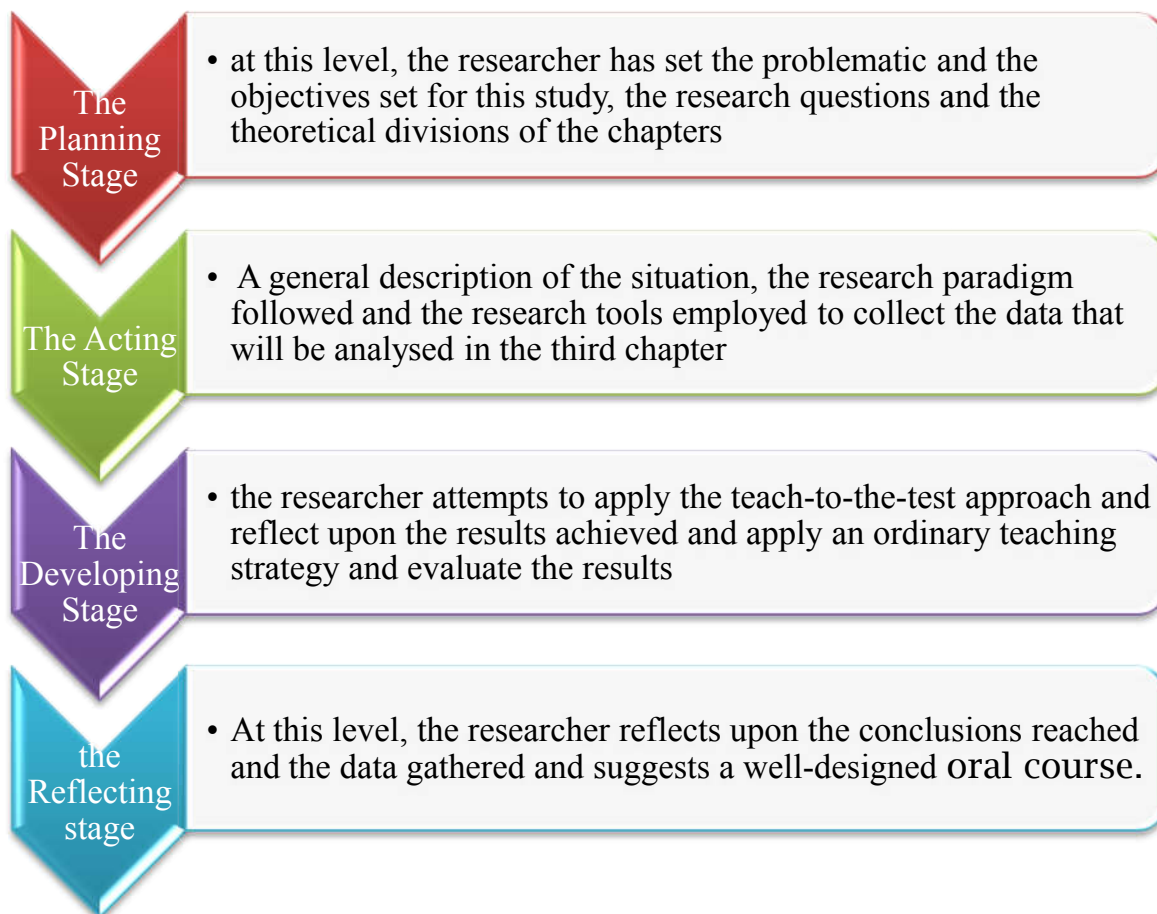


Diagram 2.2. Action Research Plan

In a more detailed manner, and based on a multi-method approach, the researcher will adapt a two cycle plan as follows:

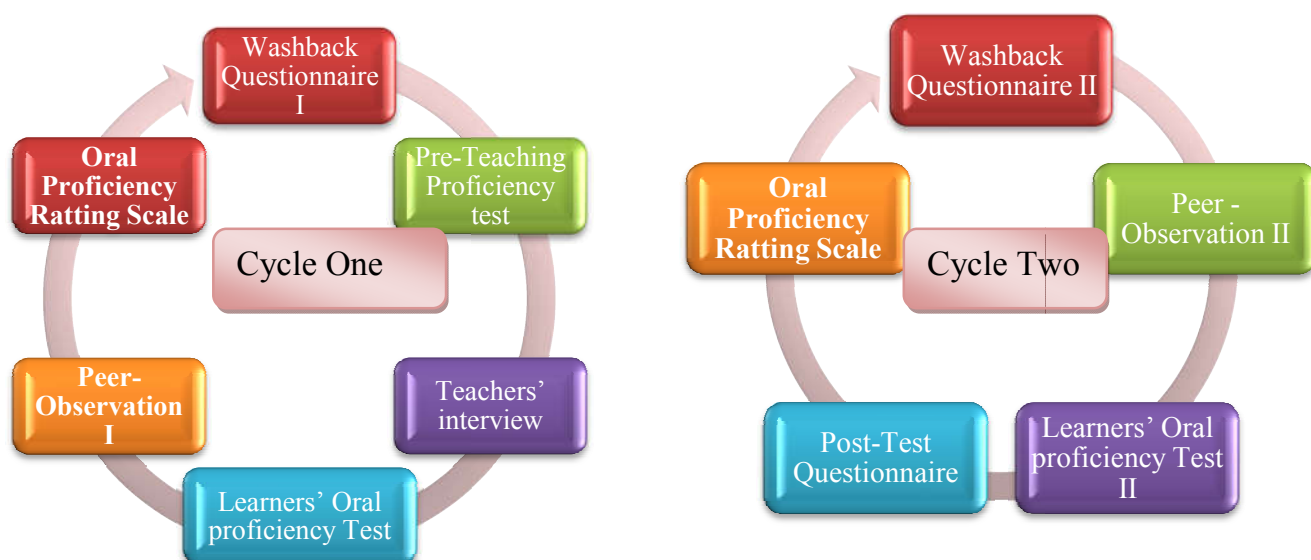


Diagram 2.3. Action Research Cycles

2.3.1. Research Roadmap, Aims and Research Questions

Assuming that washback is often deemed complex and multidimensional (Alderson and Wall, 1993), the current research work consists of a mixed method approach in designing this study. Hence, this research work attempts to demonstrate the ambiguity of washback phenomenon on the teaching strategies adopted by the teachers to develop learners' oral proficiency. It also seeks to measure the correlation between teaching-to-the-test, washback and learners' oral proficiency. In view of this, the aims of this research work are to:

- Analyze the influence of testing on teaching and its effects on learners' general oral proficiency,
- explore the influence of tests on learners' general oral proficiency,
- examine learners' and teachers' perceptions of washback effect,
- compare learners' speaking performance while applying the teach-to-the-test approach during the course of instruction and while following an ordinary teaching approach,
- seek teachers awareness of course design and test design.

The result of this research work may be helpful to teachers, educators and educationalists in general, in trying to design adequate oral courses, know how/what to test and also prepare teachers to be test designers. Hence, the study is based on the following research questions:

- How may negative/positive washback influence learners' oral proficiency?
- What is the relationship between learners' scores and the teaching method?
- What is the relationship between scores and oral proficiency?

Based on the asked questions, the following hypotheses are formulated:

- Negative washback limits learners' oral proficiency and positive washback help improve learners' oral proficiency and develop their general speaking competence.
- There is a positive statistical relationship between teaching methodology and learners' scores.

Chapter Two: An Outlook into Research Design and Methodology

- There is a positive statistical relationship between learners' oral proficiency and their scores.

In order to ease the task for the researcher, it seems wise to fix the variables of the present study as follows:

- Learners' oral proficiency is the dependent variable
- The teach-to-the-test approach and the washback phenomenon are the independent variable.

The following diagram explains this:

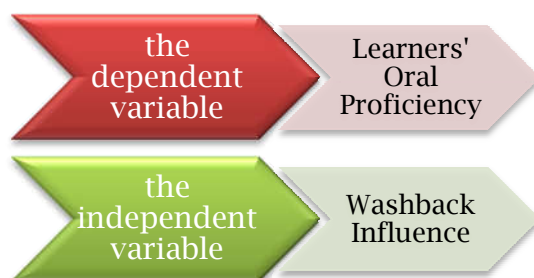


Diagram.2.4. Research Variables

2.3.2. Research Timeline and Phases

For a good time management and to clarify the research procedure, the following diagram summarizes the research timeline:



Diagram 2.5. Research Phases

2.4. Sampling Administration

To get generalizable results, the sample chosen must be representative to the whole population (Dornyei, 2003). A sample is said to be a division of the population, since it seems impractical and impossible to test each and every member of the population. The study has been conducted with second-year oral production teachers

Chapter Two: An Outlook into Research Design and Methodology

and EFL learners from Tlemcen University; English department. 35 students were randomly chosen, and seven teachers were also selected to check the matter from both sides.

2.4.1. Students' Profile

The study is concerned with second year LMD students from Tlemcen University. Students of this group, males and females, are in the age group of 18 to 26 years old. Arabic is their mother tongue, French is their first foreign language and English is their second foreign language. They are all subject of end of term examination, low and high stakes exams.

2.4.2. Teachers' Profile

The informants are seven teachers from Tlemcen University. Four holding *doctorate degree*, and one Professor holding "*doctorat d'état*" and two holding a magister degree. Their teaching experience varies from nine to thirty years, and they are in charge of the following modules: Linguistics, TEFL, Research methodology, Oral Expression, Phonetics, Linguistic Theories, Language and Culture, Communicative Language Testing and Sociolinguistics.

2.5. RESEARCH INSTRUMENTS

This section is devoted to describe the methodological procedures followed in the research work. Therefore, through the use of an action research; the researcher opts for a combination of both quantitative and qualitative data collection procedure using a triangulation process.

Several data collection tools have been used in this study, for the first cycle instrument a pre-teaching test (needs assessment test), an oral proficiency test, teachers' washback questionnaire, teachers' interview and classroom observation and an oral proficiency rating scale were used. For the second cycle, the researcher employed a learners' proficiency test, a post-test questionnaire, consisting of a Likert Scale, proficiency rating scale in addition to classroom observation. The following diagram summarizes the tools employed for the first and the second cycle.

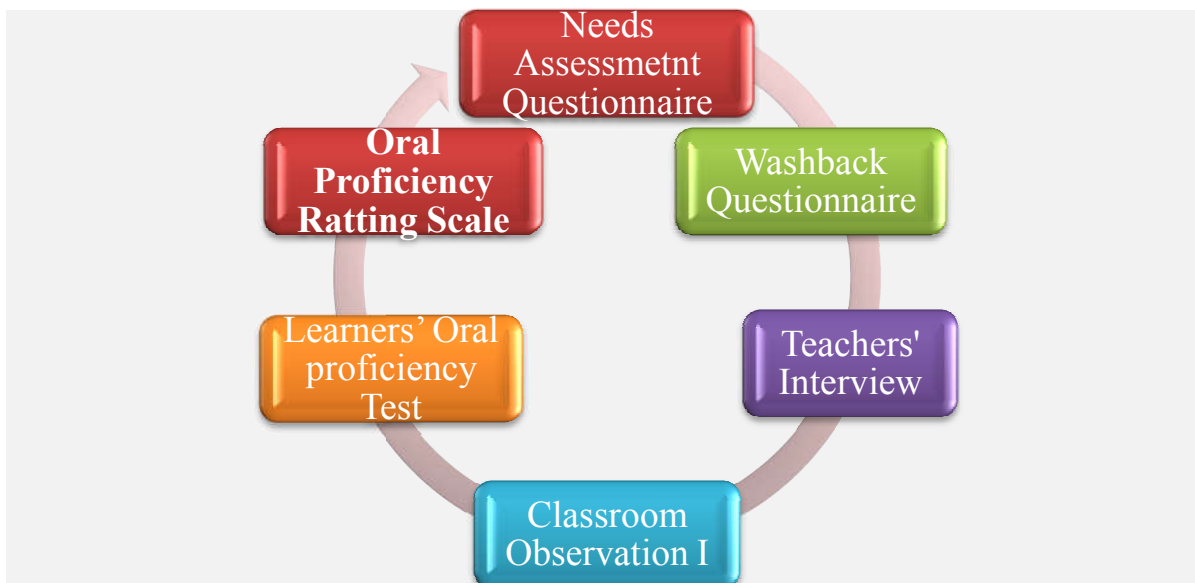


Diagram 2.6. Cycle One Research Instruments

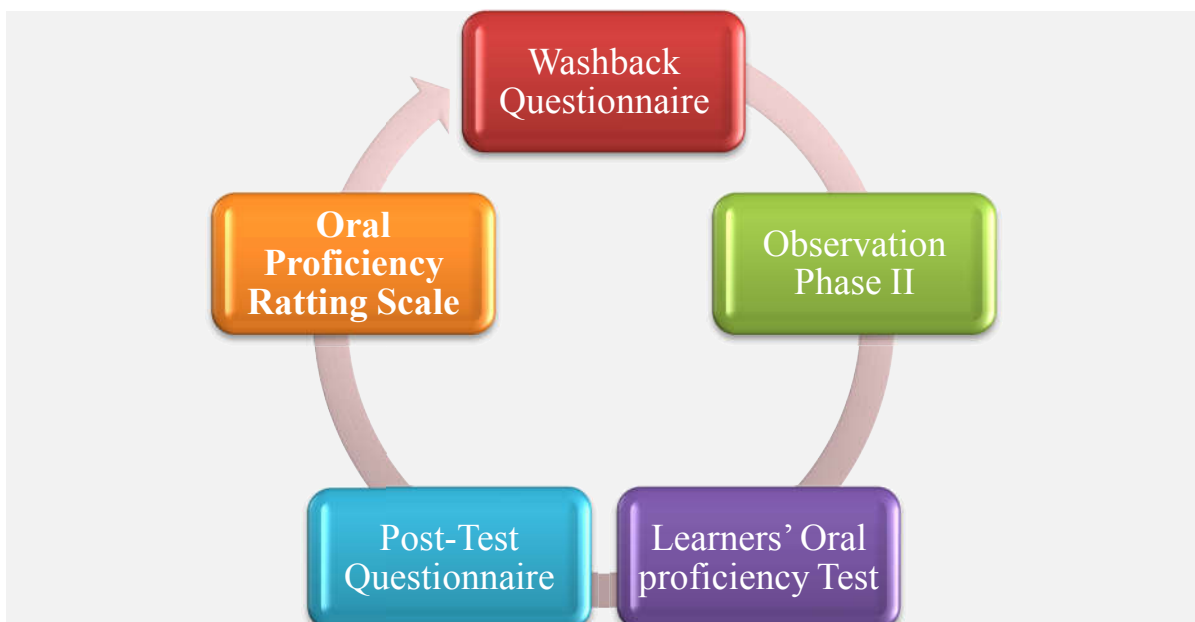


Diagram 2.7. Cycle two Research instruments

2.5.1. Cycle one

To start with, the researcher opts for the needs assessment questionnaire to gather data about learners' general proficiency.

2.5.1.1. Needs Assessment Questionnaire : Pre-Teaching Proficiency Test

It is of an immense importance in any research work to depict learners' needs and wants for an ultimate systematic course and test design. Needs analysis or needs

Chapter Two: An Outlook into Research Design and Methodology

assessment refers to the set of actions involved in gathering information about the target situation that serves as a platform for curriculum development in general and for course design in particular that meet learners' needs. Once those needs are identified they can be stated in terms of goals and objectives which, on their side, may serve as a basis for material selection, teaching strategy, test development and evaluation strategies. In this sense (1998:126) define it as follows:

First, needs analysis aims to know learners as people, as language users and as language learners. Second, needs analysis study also aims to know how language learning and skills learning can be maximized for a given learner group. Third, needs analysis study aims to know the target situations and learning environment so that data can appropriately be interpreted.

It is, thus, unquestionable that the needs analysis phase is very crucial to design, develop a language course and select the adequate teaching materials. Robinson (1991:4) proposes that needs analysis is not only just for determining the: “*what and how of a language of teaching but also as an informative database of learners, sponsors, subject-specialists and teachers in general*”.

Researchers like Dudley-Evans & St John (1998) put forward the purpose of conducting a needs assessment procedure before teaching, for instance, Brown (2001) mentions that there exist four philosophies behind the aim of conducting needs analysis in education. He (ibid cited in Stufflebeam *et. al.* (1985:13), states them as follows:

1. Discrepancy philosophy meaning the space between students' future language requirements and what they are able to do with language now.
2. Democratic philosophy referring to the needs that are preferred by the majority of the stakeholders involved in the process of language instruction.
3. Analytic philosophy – given learner characteristics and the learning processes, needs are the next things to acquire.
4. Diagnostic philosophy – like drugs for a prescription, needs are the required elements of language performance; harmful if not developed.

Chapter Two: An Outlook into Research Design and Methodology

Dudley-Evans & St John (1998: 33) have set a needs analysis model that summarizes the whole components required. It can be summarized as follows:

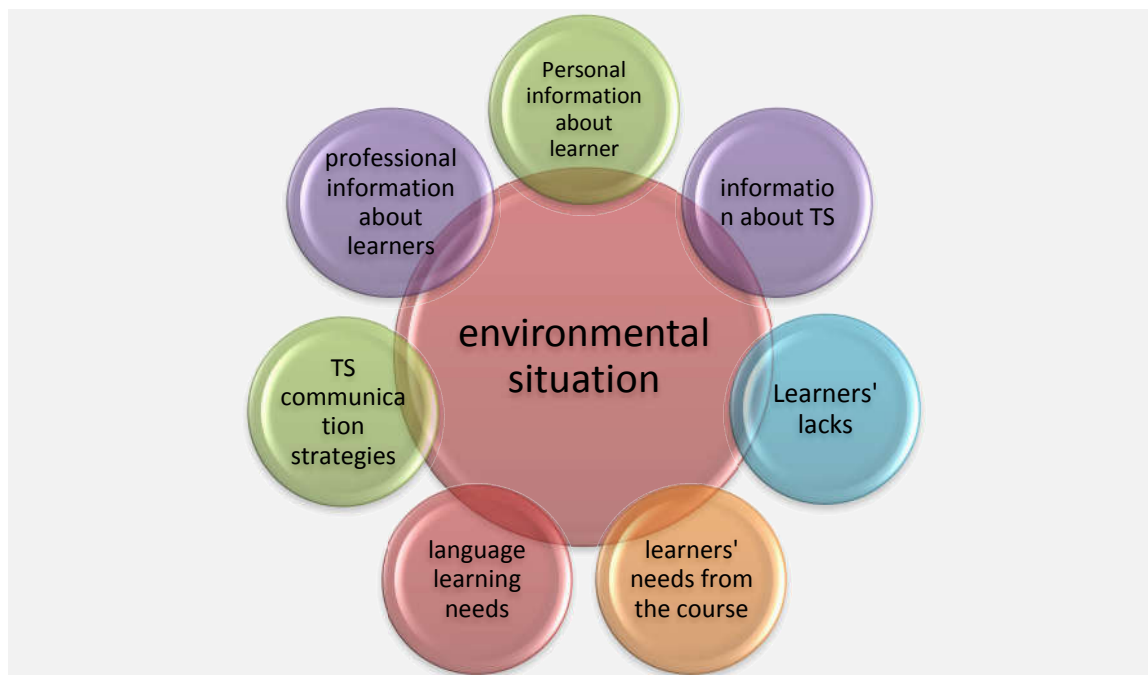


Diagram 2.8. Needs Analysis Model (adopted from Dudley-Evans & St John,1998)

This representation symbolizes one of the most comprehensive models for investigations about learners' needs and lacks. Hence, learners' needs have been identified at the onset of the investigation through a needs analysis questionnaire in order to have an idea about the learners' needs or lacks, and difficulties to help the researcher plan the teaching courses according to their levels and their needs.

In the present research work, a needs assessment questionnaire (see Appendix 'A') was administered to 35 learners in order to diagnose their strengths and weaknesses, their current level and their overall expectations about the oral course. The questionnaire was divided into two rubrics; the first rubric contains eight questions to know learners' attitudes about communicative activities performed in the oral course. The second rubric contains a Likert scale questionnaire where students are asked to tick the statement that best describes their state. Describing the questionnaire will help the researcher establish the target objectives of the research work, the objectives are summarized in the following table:

Q	Objective
1	It aims to check if learners have certain awareness about the importance of the oral course
2	this question aims to reveal learners' attitudes about the oral course in general, the aim here, is to know the attitudes of students towards the oral course and whether they liked the module or not
3	It revolves around difficulties learners encounter when performing an oral task.
4	It aims at seeking learners' awareness about oral proficiency
5	The aim of this question is to include learners' in the expected designed course, the teacher wanted to include what learners prefer as activities to make sure that they can perform activities they like
6	The aim of this question is to get an idea of what learners already did in the previous oral courses, and what was the most engaging activity for them
7	This question was put on purpose to get to know learners' clear attitudes about the oral course as well as the oral module. This question tends to raise their motivation and self esteem.
8	The last question was open to present ideas about whether they have special needs, learning difficulties, some disorders

Table 2.2. Needs Assessment Questionnaire Objectives

Following the open ended questions, the researcher provided a number of Likert questions to investigate learners' overall opinion about their current levels. The following table summarizes the objective of each statement:

Chapter Two: An Outlook into Research Design and Methodology

Statement	Objectives
I can express ideas and opinions with very high level of fluency.	The aim is to get students' points of view about their level of fluency.
I can Show very high level of comprehension, confidence and accuracy in speaking.	This statement seeks to find out whether students are aware about their current level of confidence and accuracy or not
I display high levels of critical thinking	The aim is to know whether students can have a self assessment about critical thinking or not.
I can interact effectively in speaking English	This statement would give an idea about whether students interact or not
I am able to present information in sequence and interact accurately	The aim is to know if students are able to present a talk in sequence and interact accurately or not.
I am able to understand idioms and various meanings of words	The aim is to know if students' attitudes about idiom teaching.
I am able to convey the message according to the intention.	The aim is to see if students convey the message according to their intention.

Table 2.3. Questionnaire Objectives

Thus, the coding strategy was put as follows:

Strongly agree	Agree	Strongly Disagree	Disagree	Not Sure
5	4	3	2	1

Table 2.4. Coding strategy

To be qualified as a reliable questionnaire, items within the questionnaire must be homogeneous as claimed by Dornyei (2003:68) who believes that:

It is obvious, however, that multi-item scales are only effective if the items within a scale work together in a homogeneous manner, that is, if they measure the same target area. In psychometric terms this means that each item on a scale should correlate with the other items and with the total scale score, which has been referred to as Likert's criterion of 'Internal Consistency' (Anderson, 1985).

The term *Internal Consistency* has been commonly used in language research worldwide, Cronbach (1951:323) used the terms *Internal Consistency* and

Chapter Two: An Outlook into Research Design and Methodology

Homogeneity interchangeably noting that “*an internally consistent or homogeneous test should be independent of test length*”. The internal consistency measurement determines the degree of correlation between the different items of a questionnaire or a test; it measures the reliability of the results achieved by the questionnaire.

The reliability statistics table provides the value of Cronbach’s Alpha under the following formula:

$$\alpha = \frac{N \cdot \bar{c}}{\bar{v} + (N - 1) \cdot \bar{c}}$$

Where:

- N = the number of items.
- c = average covariance between item-pairs.
- v = average variance

The internal consistency of the needs assessment questionnaire is presented as follows:

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.976	.979	7

Table 2.4.Internal consistency

The result of the internal consistency was $\alpha = .976$ which indicates a high level of consistency for the questionnaire scale, which is reliable and optimal for this research. In order to achieve well-designed questionnaire, the researcher opts for a squared multiple correlation (henceforth SMC) equation. The coefficient of the SMC measures how well items of a questionnaire are homogeneous for better construction of items. Results of the coefficient analysis are displayed in the following table:

Item-Total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Fluency	18,91	64,023	,757	,657	,983
Comprehension	19,91	53,507	,966	,955	,968
Critical	20,03	51,257	,958	,954	,970
Interaction	19,81	52,093	,966	,956	,969
Accuracy	20,13	52,694	,963	,949	,969
Idioms	20,38	58,758	,935	,884	,972
Convey	19,97	57,902	,928	,910	,972

Table 2.5. Squared Multiple Correlation

Analyzing the findings above, all questions are reliable enough for the quest. Therefore, the researcher would not remove any of the questions from the questionnaire.

2.5.1.1.1. One-Sample Kolmogorov-Smirnov Test

Assessing the normality of data is a pre-test that is performed in each statistical research, to determine whether or not the data follows a parametric distribution. One-Sample Kolmogorov-Smirnov test is used to analyze if the data comes from a Specific distribution, it is used to see if the sample comes from data that is parametric or non-parametric. Thus, results are represented bellow:

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	Df	Sig.
Fluency	,313	32	,000	,670	32	,000
Comprehension	,256	32	,000	,864	32	,001
Critical	,265	32	,000	,824	32	,000
Interaction	,222	32	,000	,844	32	,000
Accuracy	,155	32	,049	,874	32	,001
Idioms	,209	32	,001	,912	32	,013
Convey	,231	32	,000	,902	32	,007

a. Lilliefors Significance Correction

Results displayed above are from two well-known tests of normality, namely the Kolmogorov-Smirnov Test and the Shapiro-Wilk Test. Having a small sample (<

Chapter Two: An Outlook into Research Design and Methodology

50 samples), Shapiro-Wilk seem to be more appropriate and give more accurate results. One can read from the table that all statements except accuracy statement is normally distributed; because if the Shapiro-Wilk Test is below 0.05, then, the data comes from a normal distribution.

The Kolmogorov–Smirnov statistic measures the distance between the questionnaire data and the results obtained. Since accuracy statement does not follow a normal distribution, thus one needs to compare the significance (Asymp. Sig. (2-tailed) of the Kolmogorov–Smirnov test. Results are displayed below:

		fluency	Comprehensi on	Critical	interaction	Accuracy	idioms	convey
N		32	32	32	32	32	32	32
Normal Parameters ^{a,b}	Mean	4,28	3,28	3,16	3,38	3,06	2,81	3,22
	Std. Deviation	,888	1,420	1,588	1,519	1,480	1,091	1,157
	Most Extreme Differences							
	Absolute	,313	,256	,265	,222	,155	,209	,231
	Positive	,209	,160	,173	,142	,142	,209	,231
	Negative	-,313	-,256	-,265	-,222	-,155	-,143	-,175
Test Statistic		,313	,256	,265	,222	,155	,209	,231
Asymp. Sig. (2-tailed)		,000 ^c	,000 ^c	,000 ^c	,000 ^c	,004 ^c	,001 ^c	,000 ^c

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

Results show that data follow a normal distribution, thus, it is parametric. Even though the two before the last questions' significance are .004 and .001 which is less than the alpha.005 set in the beginning of the study, still it is acceptable for the study.

2.5.1.2. Washback Questionnaire

Questionnaires are one of the most common instruments used to collect data. Brown(2001:6) provides the definition of a questionnaire as being:

Any written instrument that present respondents with a series of questions or statements to which they are to react either by writing out their answers or selecting from among the existing answers.

Chapter Two: An Outlook into Research Design and Methodology

On his side, Dornyei (2003; qtd. in Djebbari (2004:129) states that: “questionnaires are certainly the most often employed data collection devices in statistical work”. Questionnaires were employed to check the teaching methods used in teaching oral proficiency, in addition to the extent of washback effects in oral courses.

The following table summarizes the main objectives of washback questionnaire:

Statement	Objectives
Examinations influence my teaching of speaking.	The aim is to see whether exams exert an impact on the teaching methodology of speaking or not.
The courses I design for my students help them get prepared for examinations.	The aim of this question is to see whether teachers take the test impact into consideration when they design the course or not.
I use hidden curriculum in teaching speaking.	The aims to find or if teachers use hidden curriculum when hey design the course or not.
I don't teach my students according to the prescribed syllabus.	The aim was to find ou whether oral teachers teach students according to the prescribed syllabus or not.
Examinations oblige me to teach selected topics.	The aim is to see whether the oral test has a negative on the teaching or not.
I feel anxious to bring good results in the oral tests.	The aim is to find out where were teachers focus oriented. Whether to scores or to gain competencies
The current course help learners improve their oral proficiency	students satisfaction with the content of the course
Teaching test-taking techniques is most important in my class	This statement aims to know whether teachers focus on teaching test taking techniques or not.
I design my oral tests according to what I have taught in the classroom	The aim was to see whether they test items that were previously tough tot not.
Students' performance on the test reflects their abilities	the aim was to seek if teachers believe that performance in the test and their scores reflect their levels.
Examinations affect my teaching process every year.	To seek teachers' awareness about washback

Table 2.8. Questionnaire objectives

Chapter Two: An Outlook into Research Design and Methodology

The internal consistency of washback questionnaire was calculated as put below:

Cronbach's Alpha	N of Items
,961	10

The result of the internal consistency was $\alpha = .961$ which indicates a high level of consistency for the questionnaire scale, which is reliable and optimal for this research.

Before starting to analyze the gathered data, the researcher needs to know whether the data is parametric or non parametric distribution. Thus, one way to know this is to see if the variable follows a normal distribution. Hence, Kolmogorov-Smirnov normality test was computed, and the results are displayed in the following table:

		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
N		7	7	7	7	7	7	7	7	7	7
Normal Parameters ^{a,b}	Mean	2,857	4,428	3,571	2,428	3,42	2,285	3,571	3,714	3,428	4,285
	Std. Deviation	1,463	,5345	1,272	1,272	1,39	1,603	1,133	1,380	1,397	,4879
	Most Extreme Differences	Absolute	,354	,360	,346	,346	,373	,285	,362	,296	,373
	Positive	,217	,360	,225	,346	,198	,285	,210	,176	,198	,435
	Negative	-,354	-,286	-,346	-,225	-,373	-,211	-,362	-,296	-,373	-,279
Test Statistic		,354	,360	,346	,346	,373	,285	,362	,296	,373	,435
Asymp. Sig. (2-tailed)		,008^c	,007^c	,011^c	,011^c	,004^c	,089^c	,006^c	,007^c	,004^c	,000^c

Chapter Two: An Outlook into Research Design and Methodology

Results demonstrate that the data collected are not normally distributed (Sig ≥ 0.05) thus, it does not follow a parametric distribution. Therefore, Kendall test should be counted. Kendall tau-b (τ_b) test is a non-parametric test that computes the strength of independence between items of the questionnaire. The following formula is used to calculate the value of Kendall:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

Nc= number of concordant

Nd= Number of discordant

Results are displayed in the following table:

Test Statistics	
N	7
Kendall's W ^a	.779
Chi-Square	36.446
Df	9
Asymp. Sig.	.000

Kendall's Coefficient of Concordance

Results displayed in the table show .779 which represents a high value of Kendall's which denotes a strong association. A high significant Kendall's coefficient is considered very good (see Kendall's Test Normes). This value denotes that all teachers are applying essentially the same variable, i.e., all of them follow the same method of teaching and testing the oral proficiency.

Poor agreement = Less than 0.20
Fair agreement = 0.21 to 0.40
Moderate agreement = 0.41 to 0.60
Good agreement = 0.61 to 0.80
Very good agreement = 0.81 to 1.00

Table 2.12. Kendall's Test Normes

2.5.1.3. Learners' Oral Proficiency Test

Learners' oral proficiency was assessed based on a test that allows the test takers to show how well they can communicate orally in the target language. There are myriad of approaches of oral proficiency assessment, including direct and indirect means. Direct tests including assessments conducted by a live face-to-face interview, or indirect tests, i.e., testing methods that rely on recordings, computer programmed tests and test booklets.

The oral proficiency test is a live interview between the tester and the test taker, it is a 10minutes conversation. It is a valid and a reliable test that measures how well a person speaks a language. The process is standard; it assesses the learners' general speaking ability, through measuring language production. The test taker will be addressed a series of personalized questions elicited and rated according to the general guidelines of oral proficiency grading scale of the ACTEFL (American Council on the Teaching of Foreign Languages).

The oral proficiency test aims at measuring how much of a language someone has learned. It is not linked to any particular course of instruction, but measures the learner's general level of language mastery (Benmostefa,2014). The testing paradigm of oral proficiency is based on high frequency-count vocabulary and general basic grammar, namely the TOEFL and the ELPT (English Language Proficiency Test) which are used to measure English language proficiency of students aiming at carrying their studies in American universities. Valette (1977:6) posits, *'the aim of a proficiency test is to determine whether this language ability corresponds to specific language requirements'*.

Once the test was administered, learners had to carry out the test and here, the real assessment begins, the recorded data were transcribed manually through the Speech Recognition Rater System (henceforth, SRRS). The researcher opts for two assessment strategies, a human rater and the SRRS rater. The aim behind including the human rater is that no system can replace the human brain, thus, the human rater evaluates the content and the viability of answers, the automatic rater would then,

Chapter Two: An Outlook into Research Design and Methodology

score according to the number of pauses, repetitions and restarts and the speech rates. The following table illustrates the main objectives of the oral proficiency test:

Part	Questions	Objectives
Part I	<ul style="list-style-type: none"> • What's eating you? • Bite the bullet, • In this neck of the woods, • The bottom line, flunk a test, • Have got it made, • Have no clue, • Beyond the shadow of a doubt . 	This question was to identify the meaning of 6 idioms from the list (done previously in class). The mark allotted to this question was 3.
Part I	<ul style="list-style-type: none"> • Tell us about a film you really like • Are you doing anything special this weekend • What acts of kindness have you performed? 	The second Question was to answer on questions of their choice about topics dealt with previously in class.
Part II	<p>Q1. Talk about the photograph (three minutes)</p> <p>Two people helping other people in different situations.</p>	The first question was to test the subjects; speaking ability. It was about describing a picture provided, the situation given was not already taught previously in the classroom, they had to describe it for three minutes.
Part II	<p>Phobias are strong fears;</p> <p>“claustrophobia” is the fear of being stuck in a small space, such as an elevator.</p> <p>Do you or anyone you know have any phobia?</p>	The second question was designed to test learners' speech fluency and their general subjects' speaking ability, learners' had five minutes to express their views. The researcher used the extempore assessment tools to record the data (see appendix).

Table 2.9. Oral Proficiency Test Objectives

2.5.1.4. Oral Proficiency Rating Scale

Oral proficiency denotes the ability to use the language in real world situations in a spontaneous non-rehearsed manner (Stacey: 2020). Learners’ oral proficiency rating scale measures how well a language learner can do regardless of how the language was learned, i.e., despite the teaching methodology of the teacher, the density of washback (high or low impact), despite the affective filters, the performance should be acceptable and appropriate to the one of a native speaker.. The oral proficiency rating determines if the learner provides sufficient evidence of all assessment criteria according the grid provided by the researcher. The following table explains the rating scores of each tests criterion.

Criteria	1	2	3	4
Comprehension	Demonstrate s poor and little Mastery of the topic Makes inaccurate definitions of concepts	Demonstrates basic knowledge about the topic Makes inaccurate description of concepts	Demonstrates adequate knowledge about the topic Makes adequate description of concepts	Demonstrates accurate and thorough knowledge about the topic Makes thorough description of concepts
Fluency	Control of pace is poor	Speaks too slowly or too quickly.	Control of speed is adequate.	Consistent control of speed.
Vocabulary	Ineffective oral communication due grammar mistakes, incoherence	Ineffective oral speech due to lack of vocabulary, and lack of mastery of language	Effective, well-organized, coherent oral communication.	Clear and effective Oral communication delivered.
Pronunciation	Uses unclear pronunciation .	Uses clear pronunciation occasionally.	Uses clear pronunciation most of the time.	Uses clear sound pronunciation all of the time.
Grammar	Speech full of grammar mistakes	Moderate use of grammar mistakes	Uses correct grammatical s entences	Uses correct sound language all the time

2.5.1.5. Teachers’ Interview

Interviews present an in-depth image about three main areas. First, to know more about learners’ oral proficiency from a teacher’s angle. Second to seek information about their teaching methodology and to seek their awareness about test impact in particular research issue. They are defined by McNamara (1999:33) as follows:

Interviews are particularly useful for getting the story behind a participant’s experiences. The interviewer can pursue in-depth information around the topic. Interviews may be useful as follow-up to certain respondents to questionnaires, e.g., to further investigate their responses.

Interviews are used in qualitative data collection; the interviewer should be knowledgeable enough about the topic under investigation to be able to conduct successfully the interview. Interviews vary according to the desired objectives; they can be structured, semi-structured and unstructured. The following table clarifies the main differences between these types:

Type	Description	Advantages	Disadvantages
Structured	Extremely rigid in their operations, . All interviewees are asked the same basic questions in the same order.	It focuses on the accuracy of different responses, it offers extremely organized data can be collected	The researcher is expected to always adhere to the list of decided questions irrespective of how interesting the conversation is turning out to be with the participants.
Semi-Structured	a guided flexible conversation where the researcher can follow any idea	Questions of semi-structured interviews are prepared before the scheduled interview, Researchers can express the	No two questions will have the exact same structure, this may cause misleading questions. Participants may question the reliability factor of these

Chapter Two: An Outlook into Research Design and Methodology

		interview questions in the format they prefer, unlike the structured interview	interviews due to the flexibility offered.
Unstructured	A conversations held with a purpose in mind, There are no guidelines for the researchers to follow and so, they can approach the participants in any ethical manner to gain as much information as they possibly can for their research topic.	Easy for researchers to try and develop a friendly rapport with the participants. This leads to gaining more insights without much effort.	As there is no structure to the interview process, researchers take time to execute these interview

Table 2.10. Interview Types (Bhat, 2015)

In the present research, an *unstructured interview* has been adopted being more adequate to the objectives settled. In such a state, the researcher will use a conversation a conversation between the researcher and the students. The interview consists of three rubrics, namely oral proficiency, course design and test design. Seven teachers will be interviewed (see Appendix D) to get their knowledge about their learners’ general oral proficiency, their teaching methods adopted and their attitudes about the current applied course. Besides, teachers will be asked about their efforts in designing a speaking course, i.e., how teachers choose materials, courses and approach according to their learners; needs. Ultimately, they will be interviewed about their test design phase, their assessment strategies used and their class time test-related instruction and practice within an oral course. The following table sums up the main objectives of the interview:

Chapter Two: An Outlook into Research Design and Methodology

Rubrics	Objectives
Rubric One	General information about teachers' experience degree and post graduate option
Rubric Two	<p>Contains four questions that revolves around information about <i>learners' oral proficiency</i> and their real experiences with their students.</p> <p>Q1: Seeks information about learners' oral proficiency level and whether it is good, fair or poor.</p> <p>Q2: the researcher seeks to know whether teachers follow any specific program or teach with a course that has been selected for them.</p> <p>Q3: the question reveals whether or not the existing course is of any help to develop learner's proficiency level or not.</p> <p>Q4: This question reveals the learners' oral proficiency needs and lacks, knowing the learners' needs will help design the course that best suits their profiles</p> <p>Q5: This question particularly revolves around whether teachers are aware of what constitute an oral proficiency training or not</p> <p>Q6: the scope of this question is to seek teachers' emphasis, if it is on, the oral proficiency, teaching methodology, course objectives or test objectives.</p>
Rubric Three	<p>This rubric surrounds teachers' perceptions and strategies in the phase of <i>course design</i>.</p> <p>Q7: seeks to see whether teachers conduct a pre-teaching needs assessment survey.</p> <p>Q8: this questions teachers' awareness of course design step and procedures.</p> <p>Q9: this question seeks to find out what are the learners' oral proficiency needs.</p> <p>Q10: this question seeks to find out what is the phase that teachers find it most challenging.</p> <p>Q11: this question tries to find out whether teachers apply the teach-to-the-test approach in preparing learners' for the test.</p> <p>Q12: This questions reveals whether teachers follow any teaching program or not.</p> <p>Q13: this question revolves around finding out whether the present course is helping learners oral proficiency.</p>

Rubric Four	Deals with the test design phase, it seeks to find out how do teachers design their tests, and whether or not those test impacts their teaching or not. Q15: this questions reveals whether teachers are aware of oral proficiency components or not. Q16: this questions seeks to see whether teachers do follow an assessment strategy or not Q17: this question seeks to find out whether they do apply <i>clone teaching</i> or not. Q18: this question revolves around what is actually happening in the class time by knowing the time devoted to test preparation. Q19: this question seeks to describe the activities performed in the classroom. Q20: this questions seeks to see whether they design their tests according to what is done previously or not. Q21: this question is designed to seek teachers needs in teaching, and what are their expectations from the university institute. Q22:the last question revolves around what negative points have been bought by the test
--------------------	--

Table 2.11. Teachers Interview Objectives

2.5.1.6. Classroom Observation-Phase I-

When embarking into observation technique to data collection, the researcher asked another teacher to come attend and fill in the observation grid designed for the study (see Appendix ‘E’). This aims at providing a general description of what happens in the classroom as objectively as possible, and without influencing the events participants are engaged in. In view of this, Mason (1996:60) notes that observation usually refers to *"methods of generating data which involve the researcher immersing [him or herself] in a research setting, and systematically observing dimensions of that setting, interactions, relationships, actions, events, and so on, within it"* . These collected data are combination of field notes, intuitions of the researcher, personal opinions, impressions, and even audio or visual recordings...etc.

Chapter Two: An Outlook into Research Design and Methodology

Hence, the researcher tends to observe and document the types of teaching practices, activities, methods teachers use, and establish the existence of washback has on her teaching practices within exams context. Thus, the observation grid focuses on aspects of the lessons related to teaching, the nature and scope of teaching activities and test preparation tasks in addition to in situ exam.

2.5.2. Cycle Two

The second cycle starts two weeks after the first oral proficiency test, where the teacher has administered the same oral proficiency test to see if students have learned for the test only, and to compare the two results. The second cycle will start after reflecting on the results of the second test, i.e., whether or not following the same teaching method or change it to reach the goals set in the beginning.

The course designed is inspired by different course books that aim at developing learners' speaking proficiency in general such as '*All Clear*', '*Face to Face*' and '*Just*' series. It is composed of many units holding various aims to develop the necessary speaking skills acquired to express their thought, ideas and feelings in various situations. The following table suggests the units that are designed in the first cycle of the action research:

Unit	Theme	Speaking & Listening	Pronunciation	Language Function & Vocabulary
Unit 1	Colors and Moods	Express your mood, Feelings and Moods Starting a new class and feelings about this pressure	English Sound System: Vowels	Language used in Formal/ informal conversations,
Unit 2	If I go to college ...	Discussing point of views on jobs and future possibilities, Accepting the other's opinions.	Stress and Intonation	Agreeing and disagreeing
Unit 3	World Famous Personalities	Speak about inventions between past and present	Voice raising/ falling	Inventions that improve communication
Unit 4	Advertising	Persuading people to buy your product,	Persuading intonation	Convincing and persuading expressions

Table 2.12. Course Design

Chapter Two: An Outlook into Research Design and Methodology

Based on the designed course, the test will be based mostly on what have been taught, i.e., learners are supposed to follow the instructions done previously to prepare for the test. This method of testing is known as ‘teach-to-the-test’. (see chapter1 for further details).

In this vein, it should be noted that the way students prepare themselves for the test depends on how they perceive the test (before, during and after the test), and these effects can have either positive or negative influences on learning. The test is based on what students have dealt with and they expect what to be asked during the test.

The 2nd Cycle, however, will be based on designing another course which encloses general teaching such as debates, news discussion, presentations, group works...etc. This course aims at developing learners’ general linguistic competence, communicative competence, knowledge construction and confidence boosting. This is to prepare competent learners to become able to speak about different topics never prepared before in the classroom. In fact, this is the main aim behind developing a general English speaking level and not preparing learners who are able to speak only about what have been taught before in the classroom.

The test in such a scenario will be opened to speak about anything selected whether political topics, current issues, general knowledge...etc. This way, learners will not be limited and will prepare themselves to speak about any topic not previously rehearsed, i.e., there will be no chance for rote learning in a speaking course.

2.5.2.1. Washback Questionnaire (II)

Among the hypotheses put forward in the present study was that tests have great impact on the teacher’s methodology and choice of content. In the first cycle, there was indeed, an observed washback resulted from the teachers’ interviews, learners’ scores and the classroom observation, this gave birth to the second cycle. Teachers feel that oral tests are one of the factors that influence their teaching, and that they are not completely free to choose whatever content they like. Thus, teachers were asked purposefully to fill in a questionnaire (see appendix G).

Chapter Two: An Outlook into Research Design and Methodology

The questionnaire will be designed to gather data about teacher's attitudes about tests, how does the test influence their teaching, their course content and their teaching methodology. The following table summarizes the objectives:

Statement	Objectives
I always analyze my students' needs before I design my oral course.	The question aims to know whether teachers conduct a needs assessment survey before they design their course or not.
I develop the teaching method with which I feel more comfortable.	To seek whether teachers adopt new teaching methods or the stick to the method that makes them feel comfortable.
I am satisfied with the teaching methodology I follow in teaching speaking	He aim is to know teachers' attitudes about he methodology they follow.
I follow the curriculum and the established syllabus.	To see if they follow the prescribed syllabus or not.
I see no importance to teach a new topic that will not be examined.	To see if teachers have an idea about negative and positive washback
My main objective is to design adequate courses that develop learners' general oral proficiency.	To see the goals set at the beginning of the course.
I teach learners to prepare them be good language speakers.	To look for teachers' awareness about the course objectives.
Examinations do not oblige me to teach selected topics.	To see if they teach exam selected topics or not.
I emphasize on teaching speaking rather than on scoring.	To inquire about teachers' emphasis and short term objectives.
I do not design my tests according to what I have taught in the classroom	The aim is to determine whether they follow a teach to the test approach or an ordinary teaching.
I do not test only what I taught, I keep learners expect everything to develop their proficiency.	The aim is to determine whether teachers use blended testing or hey stick to what was the course about.

Table 2.13. Questionnaire Objectives

Chapter Two: An Outlook into Research Design and Methodology

Pre-tests were computed to check reliability of the questionnaire items, Split Half test was used. Results are displayed below:

Cronbach's Alpha	Part 1	Value	,891
		N of Items	6 ^a
	Part 2	Value	,883
		N of Items	5 ^b
	Total N of Items		11
Correlation Between Forms			,958
Spearman-Brown	Equal Length		,978
Coefficient	Unequal Length		,979
Guttman Split-Half Coefficient			,949

a. The items are: S1, S2, S3, S4, S5, S6.

b. The items are: S6, S7, S8, S9, S10, S11.

We notice that different parts of the questionnaire statement produce two different estimation of the reliability coefficient. When questionnaire statements are split, items are highly correlated within the group of the questionnaire. Split half coefficient is close to the highest value. Since questions are highly correlated split half coefficient reached its highest values, i.e., this questionnaire is reliable.

2.5.2.2. Classroom-Observation- Phase II

After administering the washback questionnaire, the second part of this research is to expand on the rationale behind using a classroom observation. First of all, the researcher has to be sure of the current situation, so, a peer teacher will be invited to observe the teaching process and check the teaching methodology followed. This is done because the teacher tester cannot decide whether his test produce positive or negative washback. There are many factors that determine the kind of washback exerted by the test, mainly what happens in the classroom and how the course is taught, in this sense, Cheng (1997: 40) states that: “*the quality of washback effect might be independent of the quality of the test*”. Subsequently, Morrow (1986) stresses out: “*in essence an examination of washback validity would take testing researchers*

Chapter Two: An Outlook into Research Design and Methodology

into the classroom in order to observe the effect of their tests in action” (cited by Cheng, 1997: 40).

2.5.2.3. Oral Proficiency Test

Conventionally, speaking tests have been likely to consist of meetings between two people, the teacher and the student, (one-to-one testing). Oddly enough, in the second term test, the assessment format has changed, taking into account the affective variables that hindered learners’ performance in the first term, the teacher has changed the assessment paradigm.

The teacher at this level, opts for adapting Oral Proficiency Interview OPI expanded and designed by the American Council on the Teaching of Foreign Languages ACTEFL in collaboration with Education Testing Service ETS. The underlying principle behind using this test is to get a sample of speech that can be measured using the ACTEFL proficiency assessment guidelines as shown below:

Guidelines	Examples
Content	greetings, health, family, daily routines, work, study
Tense	past, present, future and expressions of agreement or disagreement
Accuracy	presence or lack of errors that may or may not interfere with the communication

Table 2.14. ACTEFL Proficiency Assessment

Assessment will be a one-to-one interview between the teacher and the student. Scoring will be based on four assessment stages: a warm up. Level check, probes, and finally wind check. They are well presented in the following tables:

Stage	Description	Duration	Purpose
I	Warm-up	Less than two minutes	Making the interviewee feel at ease. It consists of greetings, and exchanging of everyday social amenities;
II	Level-check	Five minutes	Checking the highest level of proficiency of the interviewee.
III	Probes	Three minutes	Probes are unexpected challenging questions designed to spur learners' oral proficiency. If this phase is successful then this is a good indicator of a good oral proficiency.
IV	Wind-Down	Less than two minutes	Taking into account the interviewee's level, the teacher ends the conversation by thanking him.

Table 2.15. ACTEFL Proficiency Assessment Stages

2.5.2.4. Proficiency Rating Scale

The oral proficiency rating determines if the learner provides sufficient evidence of all assessment criteria according the grid provided by the researcher or not. Thus, the following table explains the rating scores of each level criteria:

Novice-Low	Oral production consists of isolated words and perhaps a few high-frequency phrases
Novice-Mid	Insufficient vocabulary, show frequent long pauses, repetition of teacher's words, Learners are able to satisfy the necessities of a basic communication retrieving learned utterances
Intermediate-Low	Able to handle successfully a limited number of interactive, task-oriented, and social situations. espond to simple statements, and maintain face-to-face conversation, although with much linguistic inaccuracy. Interviewee can perform tasks as introducing self, ordering a meal, asking directions, and making purchases. Vocabulary is adequate to express only elementary needs. Misunderstandings may arise but with repetition, the Intermediate-Low speaker can generally be understood by interlocutors accustomed to foreigners. Able to handle successfully a variety of uncomplicated, basic,

Intermediate-Mid:	and communicative tasks and social situations. The interviewee can talk simply about self and his or her family members. Can ask and answer questions on topics beyond most immediate needs Pronunciation may continue to be strongly influenced by first language and fluency may still be strained. The Intermediate-Mid can generally be understood by interlocutors accustomed to foreigners.
Intermediate-High:	Able to handle successfully most uncomplicated communicative tasks and social situations. Can initiate, sustain, and close a general conversation with a number of strategies appropriate to a range of circumstances and topics, but errors are present. There is emerging evidence of connected discourse, particularly for simple narration and/or description. The Intermediate-High speaker can generally be understood even by interlocutors not accustomed to foreigners, but repetition may be required.

Table 2.16. ACTEFL Proficiency Assessment Grid

2.5.2.5. Post-test Questionnaire

As it was aforementioned a post test questionnaire will be administered to students. The questionnaire consists of a Likert Scale made up of a series of statements. Participants were asked to indicate the extent to which they agree or disagree on the statement; putting a tick on the right box; ranging from “*strongly agree*” to “*strongly disagree*”. The reliability coefficient of the test anxiety was .82 indicating a *high reliability*. Questions vary from positive to negative statements and the objective of each question will be displayed in the following table:

Period	Statement	Purpose
Course Design	Statement 1-7	aim at gathering information about learners attitudes about the course, test and the teaching methodology followed by the teacher psychological state <i>before</i> the test

Test Taking	Statement 8-18	The following statements revolve around students' feeling <i>during</i> the test, their psychology, stress management, levels of concentration and their physiological state. In addition, to how they prepared for the test.
Oral Proficiency	Statement 19-23	The last three statements are about <i>oral proficiency</i> , which intends to get information about the proficiency level of learners.

Table2.17.Post Test Questionnaire Objectives

To check reliability of the statements, which statement to keep which to remove, reliability tests were computed. Results are shown below:

Reliability

Scale: ALL VARIABLES

Reliability Statistics

Cronbach's Alpha	N of Items
.993	24

The result of the internal consistency was calculated $\alpha = .993$ which is reliable, excellent and optimal for this research.

In order to apply different tests on this questionnaire, the researcher has first to test and see if the gathered data follow a normal or an abnormal distribution, is it parametric or non- parametric. The table is too large containing 900 data item. Thus, results are presented in (Appendix O)

Kolmogorov-Smirnov Test and the Shapiro-Wilk Tests give more appropriate and more accurate results. One can read from the table that all statements with no excluded item statement follow a normal distribution; because if the Shapiro-Wilk Test is below 0.05, then, the data comes from a normal distribution.

Kendall tau-b (τ_b) test is a non-parametric test that computes the strength of independence between items of the questionnaire. It tends to assess the trend of agreement among all respondents.

Test Statistics

N	32
Kendall's W ^a	,511
Chi-Square	376,328
Df	23
Asymp. Sig.	,000

a. Kendall's Coefficient of Concordance

The table shows $t=511$ which represents a high value of Kendall's which denotes a strong association. A high significant Kendall's coefficient is considered very well. This value denotes that all students answered approximately in the same manner.

The mean (the measure of central tendency) and the standard deviation (the measure of the dispersive tendency) were calculated and the following table demonstrates the students' scores:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
S2	32	4,2813	,88843	,15705
S3	32	3,8438	1,27278	,22500
S4	32	3,9063	1,35264	,23912
S5	32	4,6563	,48256	,08531
S6	32	3,2813	1,78225	,31506
S7	32	3,3750	1,28891	,22785
S8	32	3,3750	1,51870	,26847
S9	32	3,8750	1,33803	,23653
S10	32	3,1875	1,20315	,21269
S11	32	3,2813	1,41955	,25094
S12	32	3,0000	1,48106	,26182
S13	32	3,1563	1,58845	,28080
S14	32	2,9063	1,30407	,23053
S15	32	3,0000	1,36783	,24180
S16	32	3,6875	1,40132	,24772
S17	32	3,0000	1,36783	,24180
S18	32	3,5938	1,45601	,25739
S19	32	4,0000	1,10716	,19572

S20	32	3,5625	1,58496	,28018
S21	32	2,5313	1,39085	,24587
S22	32	2,9063	1,51038	,26700
S23	32	3,0000	1,31982	,23331
S24	32	2,9688	1,40240	,24791
S25	32	3,5938	1,52102	,26888

Table 2.25. Post Test Questionnaire Pre-Tests

From the above table, it should be clarified that the Standard Deviation designates the way the means of the achieved scores are distributed. Thus, a low Standard Deviation S.D. means the proximity of the data scores to the item means and the high level of S.D. denotes the dispersion of data over a large scale.

The Standard Deviation interrogates the homogeneity and the heterogeneity of the items. If the S.D. is high, learners' means are far from the means of the group and thus, the group is heterogeneous, and vice versa. (See part 3.3.4.)

2.6. CONCLUSION

This chapter provided a general overview of ELT testing in Algeria, a detailed description of the research methodology followed, instruments employed in data collection phase and the sample chosen to carry out this study to reach the settled objectives. The next chapter will be devoted to the analysis of the data collected and the interpretation of the results in hope to answer the research questions asked.

Chapter Three: Data Analysis and Interpretation

3.1. Introduction.....	86
3.2. Data Analysis Procedure.....	86
3.2.1. Qualitative Data Analysis.....	87
3.2.2. Quantitative Data Analysis - SPSS –.....	88
3.3. CYCLE ONE DATA ANALYSIS.....	88
3.3.1. Needs Assessment Questionnaire Analysis.....	89
3.3.2. Washback Questionnaire Analysis.....	94
3.3.3. Learners’ Oral Proficiency Test Results.....	96
3.3.4. Teachers’ Interview Results.....	102
3.3.5. Observation Phase I Analysis.....	109
3.3.6. Oral Proficiency Rating Analysis.....	110
3.4. CYCLE TWO DATA ANALYSIS.....	114
3.4.1. Washback Questionnaire Analysis.....	114
3.4.2. Observation Analysis Phase II.....	117
3.4.3. Oral Proficiency Test Results.....	118
3.4.4. Proficiency Rating Scale.....	119
3.4.5. Post Test Questionnaire Analysis.....	122
3.5. INTERPRETATIONS OF THE MAIN RESULTS.....	133
3.6. CONCLUSION.....	138

3.1. INTRODUCTION

Based on the theoretical basics achieved in chapter two, the present chapter seeks to go through the data analysis process and the interpretation of the results obtained. Results were gathered from the teachers' questionnaires and interview, proficiency tests, and learners' questionnaires, the oral proficiency tests, classroom observation. This chapter will summarize the main results to hopefully answer the research questions of the research work settled at the onset of this study.

The methodological process of data analysis was discussed in the previous chapter, and the results of this analysis will be reported in this chapter. As discussed previously, two cycles were used to apply the action research plan; the first cycle adopts a teach-to-the-test teaching methodology, while in the second cycle an ordinary approach to teaching and testing is applied.

3.2. DATA ANALYSIS PROCEDURE

As it is generally acknowledged, data analysis procedure helps drawing conclusions from the data gathered for the sake of finding solutions to the raised research problem. In this sense, data analysis helps bringing order, meaning and structure to the collected data. In this vein, Schwandt (2007:6) describes it as:

messy, ambiguous and time-consuming, but also as a creative and fascinating process...while it does not proceed in linear fashion, it is the activity of making sense of, interpreting and theorizing data that signifies a search for general statements among categories of data

Besides, Best and Khan (2006:354) hypothesize that “*the analysis and interpretation of data represent the application of deductive and inductive logic to the research*”. Regarding the process of data analysis employed in coding measuring and analyzing data, the researcher will rely on both qualitative and quantitative data analysis. Thus, the integration of quantitative and qualitative data in one research work has great potential to strengthen the research work and enrich the analysis of the findings.

On the way of making sense of the data gathered from the research instruments, the researcher employed a mixed method i.e., “*mixing or combining quantitative and*

qualitative research techniques, methods, approaches, concepts or language into a single study” Johnson (2004: 17). Mixed analysis involves the use of both quantitative and qualitative analytical techniques within the same framework, which is guided either a priori, a posteriori, or iteratively (representing analytical decisions that occur both prior to the study and during the study). Qualitative data was analyzed in a textual form and quantitative data was transformed into numerical forms using SPSS version 22 spreadsheet.

3.2.1. Qualitative Data Analysis

Qualitative data analysis is the process through which the researcher makes sense from the participants’ points of view, opinions of situations, corresponding patterns, themes, categories and regular similarities (Cohen, 2007). A good working definition seems to be that of Nieuwenhuis (2007:100) who states that: “*..qualitative data analysis tends to be an ongoing and iterative process, implying that data collection, processing, analysis and reporting are intertwined, and not necessarily a successive process*”. Besides, Gibbs (2007) points out that qualitative data analysis is a process of transformation of collected qualitative data, done by means of analytic procedures, into a clear, understandable, insightful, trustworthy and even original analysis.

Thus, qualitative data analysis is the process by which the researcher searches for general statements about relationships among categories of data. Denzin and Lincoln (2000: 3) offer the following definition: “*qualitative analysis is a situated activity that locates the observer in the world. It consists of a set of interpretive, material practices that makes the world visible*”.

Hence, the crucial goal for employing qualitative research into this research work is to discover and describe phenomenon not previously described and to view the fact from your participants’ angle. A variety of different practices are used in the collection of qualitative data, each technique with its advantages and disadvantages. Among the most commonly used qualitative data collection instruments are: questionnaires, diaries and journals, interviews, case studies and observational techniques. In this research, questionnaires (pre and post teaching), classroom

observation, and interviews will be used. Among the disadvantages of using qualitative research is that the results obtained cannot be extended to a wider population and results cannot be *less generalisable* shohamy(1989) because the findings are not tested to discover whether they are statistically significant or not.

3.2.2. Quantitative Data Analysis - SPSS –

Having designed, administered, and collected data from the target sample is '*half the battle*' Djebbari (2014:162). Thus, the second remaining part deals with data coding and processing which is "... a *process of entering data, naming and defining variables, making sure that the entry process is quality controlled, and cleaning the data to prepare for a quantitative data analysis*" LeCompte & Schensul, (1999:119). The objective of this part is to determine if there is any development in learners' oral proficiency from the teach-to-the-test approach to teaching and testing,

To submit accurate findings, the data entered into the data file should be very accurate prior to any analysis (Tabachnick, Fidell, & Osterlind, 2001). Thus, before diving into data analysis, some data coding procedures should be followed. Therefore, the quantitative data analysis phase relies heavily on numerical analysis which was performed using the *Statistical Package for the Social Sciences* (henceforth SPSS) version 22. The Alpha level was set on .05, Pearson correlation coefficient was reported throughout Pearson's Chi-Squared (χ^2). As for the nominal variables, Cramer's Phi (ϕ_c) was computed to measure whether there exist a statistical significance between teachers' teaching strategies and learners' oral performance. For the correlation analysis, Spearman's correlation coefficient was computed in order to set the ordinal data measurable.

3.3. CYCLE ONE DATA ANALYSIS

Data gathered from cycle one within the action research plan will be analyzed in this section. Results are taken from pre-teaching questionnaire, washback questionnaire, learners' oral proficiency test, teachers' interview, observation, and the oral proficiency rating scale.

3.3.1. Needs Assessment Questionnaire Analysis

This questionnaire was set to diagnose learners' strengths and weaknesses, their current level and their overall expectations about the oral course. Results are displayed below:

Results from the first question were diverse, learners have a great awareness about the importance of the oral course, they claimed that oral course help them widen their communicative competence and communicative efficiency. Some learners believe that the oral module will help them become good public speakers; others want to develop their speaking competence in terms of fluency and pronunciation.

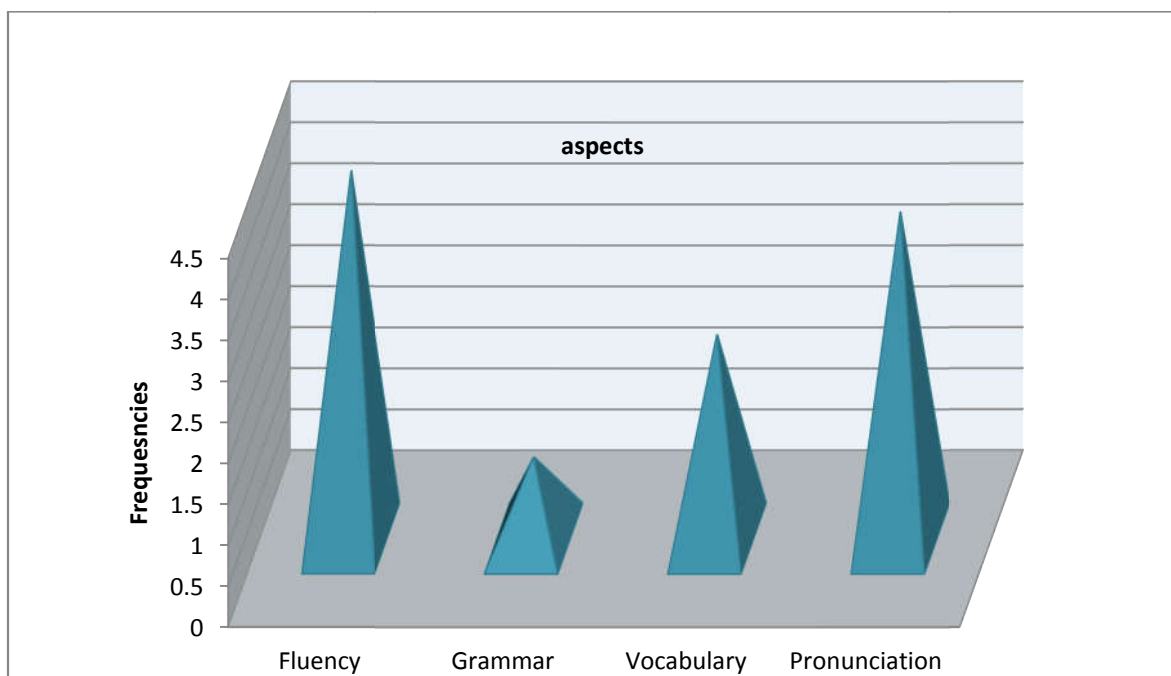
The second question revealed that each learner had a different view each according to his he psychological state. Most learners come to the oral class ready and willing to learn, the majority of students (70%) recall a good experience at the oral course; they have learned many idioms, vocabulary, phrasal verbs that they still use them, while the remaining students 30% claimed that they had a bad experience in the oral course due to many factors namely, shyness and lack of motivation, fear of public speaking due to the lack of participation in the class. When asked about the reasons behind such psychological state, they maintain that it is mainly due to the lack of participation, stressful classroom environment and sometimes due to teachers' anxiety providing behaviors.

The third question showed the difficulties learners encounter when performing an oral task. From all the answers, difficulties can be divided into two categories: linguistic and non linguistic problems as follows:

Linguistic Problems	Non-Linguistic Problems
Lack Grammar knowledge	not being brave to speak
Difficult to express words or sentence	being unconfident to speak
Lack of vocabulary	being afraid of making errors
Poor Pronunciation	being too shy to speak
	Teachers' Anxiety Providing Behaviour

Table 3.1. Speaking Problems faced by Learners

The fourth question aims at seeking learners' awareness about oral proficiency, results reveal that learners put a great emphasis on developing fluency and pronunciation at the expense of grammar and vocabulary in the oral course, results are displayed below:



Bar-Graph 3.1. Learners' Oral Proficiency Component Awareness

The fifth question aims at including learners' in the expected designed course. Answers revealed that students prefer songs, stories, dialogues, role plays and language games. They believe that these activities make learners better motivated to learn.

The sixth question results diverse answers, all learners recall their contribution in the role plays and said that it was a memorable experience that they will always keep in their memories.

The seventh question tends to raise their motivation and self esteem. Answers to this question were directed into setting goals for the two academic terms. Three students had to share their personal experience with a teacher who used to have immediate corrections to their grammar mistakes and they asked the teacher to treat their mistakes gently to avoid recalling the uneasy feeling with the other teacher.

The last question was open to present ideas about whether they have special needs, learning difficulties, some disorders. Answers were quite good, as no one really

has a language disorder or a learning difficulty, except for some who presented anxiety as being of a real threat to them.

Following the open ended questions, the researcher provided a number of questions under a Likert scale to investigate learners' overall opinion about their current levels. Through the use of the statistical tool SPSS, Cronbach alpha was computed and the results of the internal consistency was $\alpha = 0,97$ which is reliable and optimal for the research.

S1: I can express ideas and opinion with a high level of fluency

For statement one, the mean and standard deviation were calculated and displayed in the following table:

	N	Minimum	Mode	Mean	Std. Deviation
Fluency	35	1,00	5,00	3,4667	1,59164
Valid N (listwise)	35				

From the above table, it should be clarified that the Standard Deviation (SD) designates the way the means of the achieved scores are distributed. Thus, a low S.D. means the proximity of the data scores to the item means and the high level of S.D. denotes the dispersion of data over a large scale. The Standard Deviation interrogates the homogeneity and the heterogeneity of the items. If the S.D. is high, learners' means are far from the means of the group and thus, the group is heterogeneous, and vice versa. In this statement, the standard deviation was 1.59 and the mean was 3,46, the mode was 5, which denotes the most spread score. Therefore, this denotes a low estimation of the statement and thus, students still doubt about their fluency in oral performances.

S2: I can Show very high level of comprehension, confidence and accuracy in speaking.

	N	Sum	Mean	Std. Deviation
Confidence	35	108,00	3,6000	1,40443
Valid N (listwise)	35			

In this statement, the standard deviation was 1.40 and the mean was 3.60, the mode was 4, which denotes the most spread score. Therefore, this denotes a low estimation of the statement and thus, students have less self confidence in speaking. Teachers, then, need to raise their confidence to push them interact in the target language.

S3: I display high levels of critical thinking

	N	Sum	Mean	Std. Deviation
critical Thinking	35	3,00	2,4333	1,52414
Valid N (listwise)	35			

In this statement, the standard deviation was 1.52 and the mean was 2,43, the mode was 3, which denotes the most spread score. Therefore, this denotes a low estimation of the statement and thus, students have fewer levels of critical thinking skills in speaking.

S4: I can interact effectively in speaking English

	N	Sum	Mean	Std. Deviation
Interaction	35	4,00	3,4000	1,40443
Valid N (listwise)	35			

In this statement, the standard deviation was 1.40 and the mean was 3.40, the mode was 4, which denotes the most spread score. Therefore, this denotes a low estimation of the statement and thus, students still have problems in using English as a

means of interaction, they face anxiety and lack of self confidence when asked to interact in English.

S5: I am able to present information in sequence and interact accurately

	N	Sum	Mean	Std. Deviation
Presentations	35	3,00	3,0333	1,42595
Valid N (listwise)	35			

In this statement, the standard deviation was 1.42 and the mean was 3.03, the mode was 3, which denotes the most spread score. Therefore, this denotes a low estimation of the statement. Students claim that they are still unable to have a non-stop talk accurately.

S6: I am able to understand idioms and various meanings of words

	N	Sum	Mean	Std. Deviation
Idioms	35	3,00	3,4000	1,45270
Valid N (listwise)	35			

In this statement, the standard deviation was 1.45 and the mean was 3.40, the mode was 3, which denotes the most spread score. Therefore, this denotes a low estimation of the statement. Here students do not have the capacity to understand idioms and various words. Thus, this denotes that the majority of students have problems understanding idioms and various meanings of words. They simply have negative attitudes about teaching idioms in the oral course. Consequently, the teacher should be well aware of this and take it into consideration in designing the syllabus of the academic year.

S7: I am able to convey the message according to the intention.

	N	Minimum	Maximum	Mean	Std. Deviation
convey message	35	1,00	5,00	2,6667	1,44636
Valid N (listwise)	35				

Here, the standard deviation was 1.44 and the mean was 2.66, the mode was 3, which denotes the most spread score. Therefore, this denotes an average estimation of the statement. Thus, students are still unable to express their ideas freely.

3.3.2. Washback Questionnaire Analysis:

After the questionnaire distribution, each statement was coded. The scores put to each statement were entered to SPSS spreadsheets. The questionnaire was designed to gather data about teacher's attitudes about tests, how does the test influence their teaching. Results are displayed bellow:

N	Statement	Mean	St. Dev	Skewness
1	Examinations influence my teaching of speaking.	2,8571	1,46385	-,556
2	The courses I design for my students help them get prepared for examinations.	4,4286	,53452	,374
3	I use hidden curriculum in teaching speaking.	3,5714	1,27242	-1,581
4	I don't teach my students according to the prescribed syllabus.	2,4286	1,27242	1,581
5	Examinations oblige me to teach selected topics.	3,4286	1,39728	-1,079
6	I feel anxious to bring good results in the oral tests.	2,2857	1,60357	1,053
7	The current course help learners improve their oral proficiency	3,5714	1,13389	-,725
8	Teaching test-taking techniques is most important in my class	3,7143	1,38013	-1,424
9	I design my oral tests according to what I have taught in the classroom	3,4286	1,38013	-1,079
10	Students' performance on the test reflects their abilities	3,4286	1,39728	-1,079
11	Examinations affect my teaching process every year.	4,2857	,48795	1,230

Table 3.2. Teachers' Washback Questionnaire I

The mean, which denotes the most widely used measure of central tendency, or what is frequently called the average, is sensitive to a large value. If one views the results displayed in the table, it is clear that the mean is between]2,86; 4,43[. Comparing it to alpha set in the beginning of the study which goes around 5, it denotes that the sum of means is close to alpha, i.e., all teachers tend to teach oral expression following the same teaching method, i.e., the test has an impact on teachers' strategies. As for the standard deviation measure, which is the square roots of variance, is put under the equation:

$$SD = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

In this case, the structure of the standard deviation is between the two extreme values].534; 1,46[which contributes most to the sum of means. Thus, the spread set of data in this questionnaire is large. Therefore, data of this questionnaire is most spread out. As for Skewness measures, this test views the degree and direction of data asymmetry. A positive value of skewness denotes a normal distribution of data and vice versa.

One sample t - Test was computed to compare means and to test whether a population mean is significantly different from the sum of the hypothesized value or not. Results are displayed below, knowing that:

- *T Statistics* of this t-test is t=5,164
- *Df*: id the degree of freedom for the test. Df=n-1= 7-1= 6, df=6
- *Sig. (2-tailed)* is the p- value adapted to this study.

One-Sample t-Test

	Test Value = 7					
	T	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
EXAM	5,164	6	,002	2,85714	1,5033	4,2110

Course	21,920	6	,000	4,42857	3,9342	4,9229
curriculum	7,426	6	,000	3,57143	2,3946	4,7482
prescribed	5,050	6	,002	2,42857	1,2518	3,6054
Slected	6,492	6	,001	3,42857	2,1363	4,7208
Anxious	3,771	6	,009	2,28571	,8027	3,7688
Test	8,333	6	,000	3,57143	2,5228	4,6201
Totest	7,120	6	,000	3,71429	2,4379	4,9907
Abilities	6,492	6	,001	3,42857	2,1363	4,7208
Affect	23,238	6	,000	4,28571	3,8344	4,7370

- *Mean Difference* denotes the difference between the observed sample mean and the expected mean put for this study.
- *95% Confidence Interval of the Difference* is the interval for the difference between t test and the questionnaire mean.

Kendall tau-b (τ_b) test is a non-parametric test that computes the strength of independence between items of the questionnaire. Results are displayed in the following table:

N	7
Kendall's W ^a	,779
Chi-Square	36,446
Df	9
Asymp. Sig.	,000

a. Kendall's Coefficient of Concordance

Results displayed in the table show .779 which represents a high rate of Kendall's which denotes a strong association. A high significant Kendall's coefficient is considered very good according to tau-b (τ_b) value grid.

This value denotes that all teachers are applying essentially the same variable, i.e., all of them follow the same method of teaching and testing the oral proficiency.

3.3.3. Learners' Oral Proficiency Test Results

Once the test was administered, learners had to complete the test and here, the real assessment begins, as the recorded data were transcribed manually through the

Speech Recognition Rater System, henceforth, SRRS. The researcher, then, opts for two assessment strategies, a human rater and the SRRS rater. The aim behind including the human rater is that no system can replace the human brain, thus, the human rater evaluates the content and the viability of answers, the automatic rater would then, score according to the number of grammar mistakes, pauses, repetitions and restarts of the speech rates.

The internal consistency was measured to determine the degree of correlation between the different items of the test. It measures the reliability of the results achieved by the test, and whether several items tend to measure the same product and get similar results.

Cronbach's Alpha ^a	N of Items
$\alpha=0,71$	6

Table3.2.Internal Consistency.

The result of the internal consistency was $\alpha = 0,71$ which is reliable and adequate for this research.

- **The Human Rater**

After the administration of the test, the teacher was asked to give each student an overall impression mark about their performance in the recorded tapes. After giving the scores, learners' scores varied from 08/20 to 16/20. The mean (the measure of central tendency) and the standard deviation (the measure of the dispersive tendency) were calculated and the following table demonstrates the students' scores:

	N	Minimum	Maximum	Mean	Std. Deviation
score1	35	8,00	16,00	12,2031	1,95868
Valid N (listwise)	35				

The lowest score was Min=8,00 and the highest score was max=16. The overall mean of students was 12.02 which denotes an average central tendency score. As for

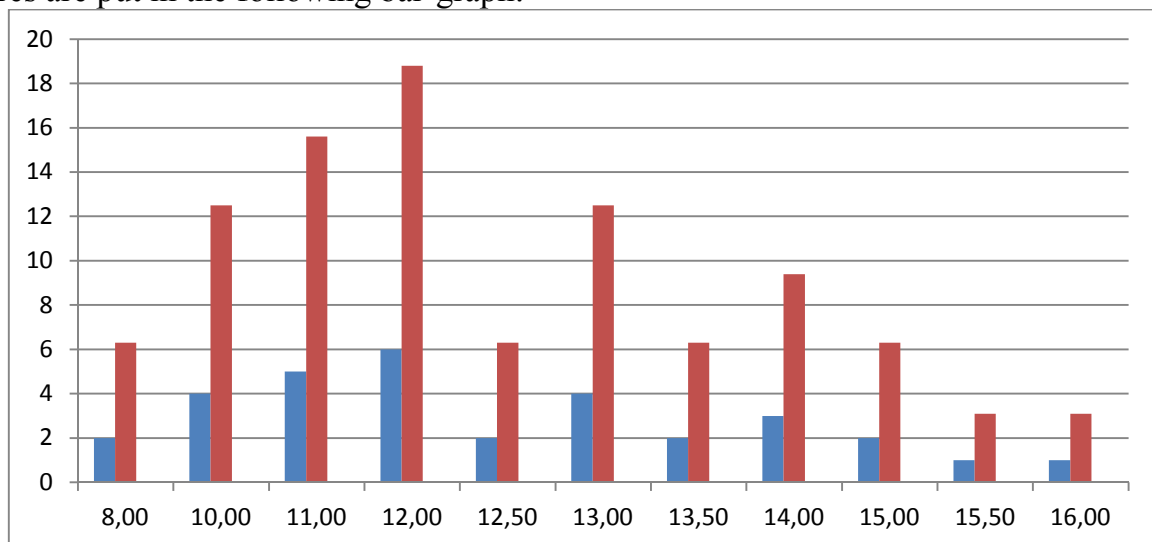
the Standard Deviation, the dispersive tendency, it was 1.95 which denotes a large amount of variation in oral proficiency scores.

Subsequently, the researcher analyzed the frequencies of scores in the oral proficiency test, results are displayed below:

Score1

Scores		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	8,00	2	6,3	6,3	6,3
	10,00	4	12,5	12,5	18,8
	11,00	5	15,6	15,6	34,4
	12,00	6	18,8	18,8	53,1
	12,50	2	6,3	6,3	59,4
	13,00	4	12,5	12,5	71,9
	13,50	2	6,3	6,3	78,1
	14,00	3	9,4	9,4	87,5
	15,00	2	6,3	6,3	93,8
	15,50	1	3,1	3,1	96,9
	16,00	1	3,1	3,1	100,0
	Total	35	100,0	100,0	

As it is displayed in the table above, the highest percentage 18,8 % was in score 12, which denotes an average score of oral proficiency. Results of the human rater scores are put in the following bar-graph:



Bar-Graph 3.2.Human Rater Scores

Only four students scored excellent results, their clarity comprehension and fluency was remarkable. Seventeen students had acceptable to somehow satisfactory results, however, eleven students did poor in the test, they were anxious shy and apprehensive throughout the test.

- **The Automatic Rater:**

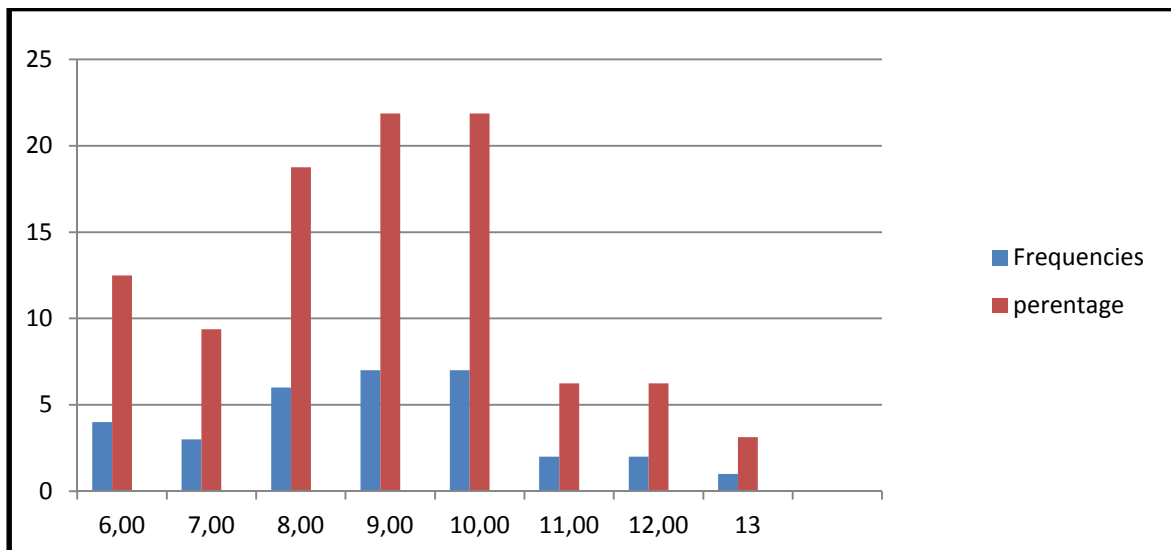
After having the human rater scores, the same tapes were put into the SRRS platform to check the system assessment. Assessment was based on a set of variables like speed, number of pauses, correct pronunciation, fluency and overall correct English grammar. Learners' scores varied from 04/20 to 13/20. The mean (the measure of central tendency) and the standard deviation (the measure of the dispersive tendency) were calculated and the following table demonstrates the students' scores:

	N	Minimum	Maximum	Mean	Std. Deviation
score2	35	4,00	13,00	8,9063	1,82030
Valid N (listwise)	35				

The lowest score was Min= 4.00 and the highest score was max=13. The overall mean of students was 8.90 which denotes a low central tendency score. As for the Standard Deviation, the dispersive tendency, it was 1.82 which denotes a large amount of variation in oral proficiency scores. Consequently, the researcher analyzed the frequencies of scores in the oral proficiency test, results are displayed below:

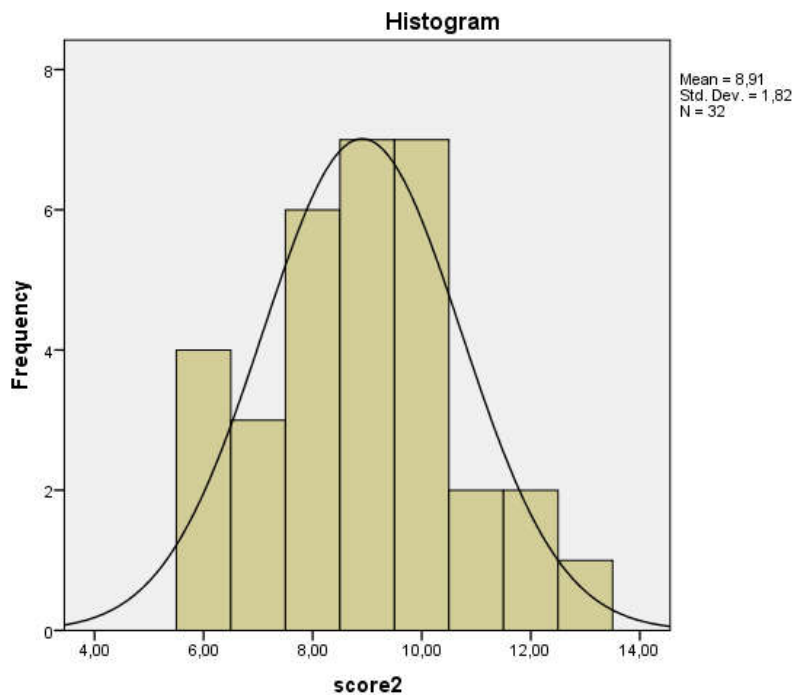
Automatic Rater	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 4,00	1	3,1	3,1	100,0
6,00	4	12,5	12,5	12,5
7,00	3	9,4	9,4	21,9
8,00	6	18,8	18,8	40,6
9,00	7	21,9	21,9	62,5
10,00	7	21,9	21,9	84,4
11,00	2	6,3	6,3	90,6
12,00	2	6,3	6,3	96,9
13,00	1	3,1	3,1	100,0
Total	35	100,0	100,0	

As it is demonstrated in the table above, the highest percentage 21.9 % was in score 9 and 10, which reflects a poor performance. Results of the automatic rater are summarized below:



Bar-Graph 3.3. Automatic Rater Scores

The frequency curve is displayed below:



Bar graph 3.4. Automatic Rater Scores Curve

- **Paired Sample T-Test:** is a test that measures if the means of the two scoring techniques (Human Rater and Automatic Rater) are equal in some situation. Results are displayed below:

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	HumanR	12,2031	35	1,95868	,34625
	AutoR	8,9063	35	1,82030	,32179

The first table presents the descriptive statistics, $N=35$ thus, the test variables are complete with no missing value. Oddly enough, as notice the mean of the automatic rater is lower than the mean of the human rater. The value of central tendency of both measures is demonstrated below:

		N	Correlation	Sig.
Pair 1	HumanR & AutoR	35	,842	,000

Knowing that the main purpose behind t -test is to know whether there is a difference in means between the two variables or not, SPSS also counts the paired samples correlation to know how associated the variables are with one another, $r=,842$ which denotes that the two scores are positively significant. P-value is 0 thus; data of this test is positively significant.

		Paired Differences				T	Df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	HumanR – AutoR	3,29688	1,06906	,18899	2,91144	3,68231	17,445	31	,000

Results of the table explain that:

- **Mean:** the average difference between the two scores,
- **Standard deviation:** the degree of deviation of the difference scores
- **Standard error mean:** this follows the equation: standard deviation divided by the square root of sample number $N=32$)
- **t:** The test statistic (denoted t) for the paired T test.

- **df:** The degrees of freedom set in this test
- **Sig. (2-tailed):** The p -value corresponding to the results of the test t and with the freedom df .

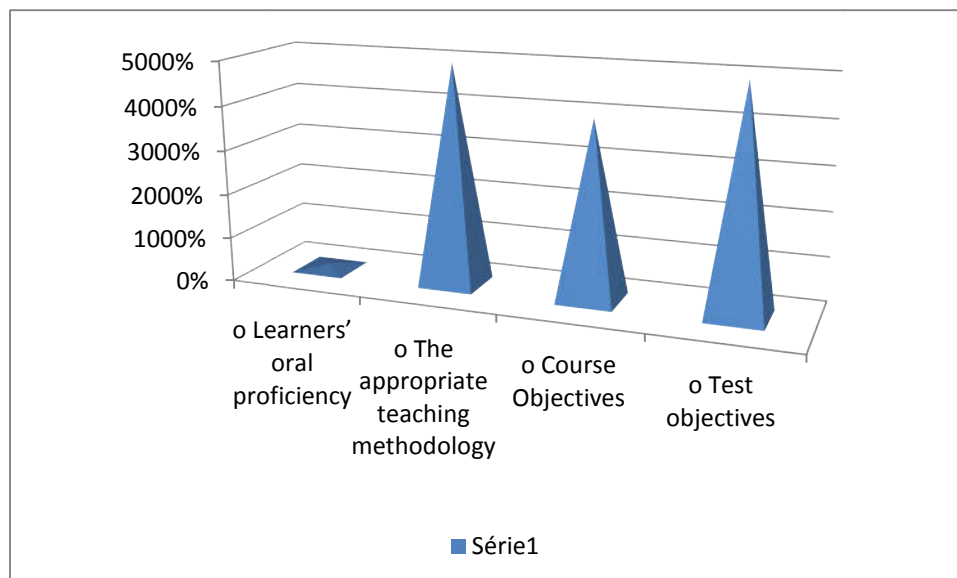
Results of this test show that the automatic rater scores are statistically insignificant when compared with the teacher's scores.

- Human rater scores and automatic rater scores are strongly positively correlated ($r = .842; p < 0.001$).
- There is a significant high difference between the human rater scores and the automatic rater scores ($t = 17.445, p < 0.001$).
- At a higher scale, the human rater scores are 3,68231 and the automatic rater scores 2,91144

3.3.4. Teachers' Interview Results

Regarding the interview, and as stated earlier in the previous chapter, it provides in-depth information about the target research issue. It was conducted with seven teachers in order to bring out the required data in this study (see appendix 'D'). It was mainly employed to further examine the research questions and hypotheses and gather significant data about the teachers' beliefs, methodologies and techniques. It spotlighted the respondents' testing techniques when it comes to oral proficiency level, their techniques in designing oral courses and their assessment strategies. The interview was divided into different parts, namely oral proficiency, course design and test design.

The first rubric revolves around gathering information about learners' oral proficiency. The first question reveals that the majority of teachers' (70%) believe that L2 learners have an average oral proficiency level, others believe that they have a good level, however 10% state that they have a poor level as displayed below:



Bar-Graph 3.5. Teachers' Beliefs about Learners' Oral Proficiency Level

The second question seeks information about teachers' followed strategies in designing oral courses. Results revealed that 30% followed selected coursebooks such as All Clear, Just Series, and Face to Face in designing their courses. However, the remaining teachers state that they use random activities such as dialogues, presentations and discussions in their courses.

When asked whether their programme is developing learners' oral proficiency, teachers state that it depends on learners' results. One teacher asserts that from the results achieved he detects his success or failure and he tries to fix the required skills. Another teacher states that his designed programme helps learners develop a certain level of proficiency and he confirms this from the results achieved in third year. Oddly enough, one teacher strongly believes that since no guidelines are provided, then each teacher is free in his selections and his teaching.

The fourth question tried to highlight learners' oral proficiency lacks. Results were at variance, they are summarised as follows:

- Lack of vocabulary,
- lack of confidence to speak,
- lack of motivation
- lack of speaking skills,
- problems of accuracy,

- little opportunities to speak English,
- lack of practice outside the classroom,
- no enthusiasm in the speaking course,
- afraid of negative evaluation,
- afraid of making mistakes,
- lack of linguistic knowledge,

As for the fifth question, all teachers state they train their learners' oral proficiency. However, when asked to justify their selection, it was found that they seem to be unaware about the main components of oral proficiency training. They view oral proficiency solely in non-academic manner, i.e., they summarise it in good speaking in general, a speech free of mistakes.

The sixth question was direct to see teachers' main focus when teaching. Results revealed that all teacher focus on raising their learners' oral proficiency as their first top priority. Besides, 4 teachers focus more on test and course objectives, and this will shape their teaching methodology.

The second rubric focused mainly on course design, and teachers' efforts as course developers. The seventh question demonstrated that the 1st step teachers follow in designing an oral course is by considering their needs and preferences. For instance, one teacher states that he updates his courses every semester to meet his learners' needs. Another teacher asserts that he administers a questionnaire seeking learners' needs. Teachers strategies used at the onset of their course design are put as follows:

- check learners' general level,
- generate an updated course,
- test learners at the onset of the year,
- administer a questionnaire,

The eighth question attempts at considering teachers' criteria selection prior the design phase. The following criteria are suggested:

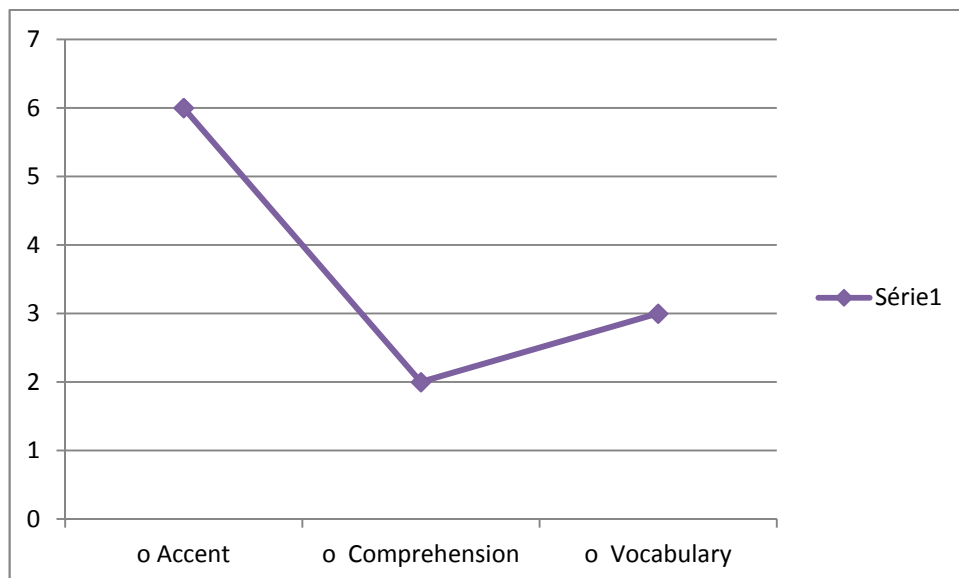
- Articulate course goal,
- Consider who are the target learners,
- Consider what do they need,

- Consider how to deliver the course,
- Select the materials needed
- Design the needed activities and assignments.

The ninth question was addressed in relation to the previous ones. The researcher attempted at confirming the previous asked questions by asking teachers' awareness of their learners' speaking needs. Results were at variance, they are put as follows:

- Learners need to be competent speakers,
- Learners need to accumulate pertinent vocabulary,
- Learners should be trained to speak fluently,
- Learners need to develop language awareness,
- Learners need to develop their confidence when speaking,
- Learners need to be able to express themselves in different topics,

Question ten asked about teachers' most challenging phases, results are displayed in the following graph:



Graph 3.6. Teacher's Most Challenging Phases

In fact, results demonstrate that all teachers start from course design then focus on test design, and finally move to scoring strategies. They justified their selection as follows;

- Course design reflects automatically test and scoring phase,
- The three phases represent a ‘*cause and effect*’ phenomenon, after designing the course, teachers should focus on testing their learners’ outcomes according to the designed course, and this will be subsequently followed by scoring their results.
- Scoring phase comes as a final step after designing and testing phases,
- The testing phase depends on the course design phase, i.e., what is taught will be tested and scored,
- The most challenging phase is the scoring phase because it is time consuming and effort demanding,
- Testing learners is always a dependent variable, and scoring is a result of the designing and testing phases.

The eleventh question focused on teachers’ use of test-related materials and their priorities in selecting test-materials. Results revealed that teachers rely on preparing students to score well in the speaking test through a measurement-driven instruction.

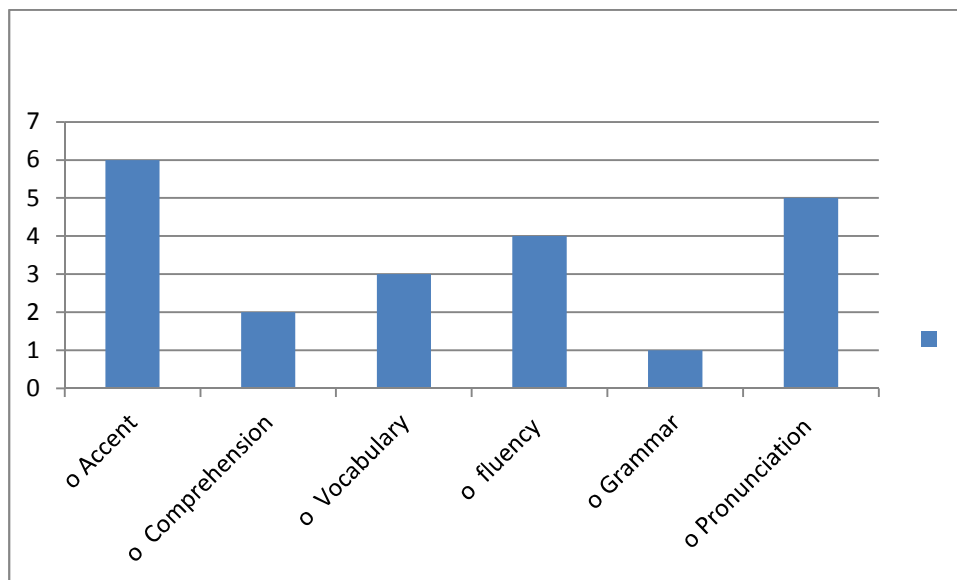
The twelfth question seeks looking for teachers’ followed programmes. Results were very poles apart, especially that every teacher follows his own designed programme. One teacher believes that a unified oral course is quite impossible since each group of students need specific skills to be developed, and he justified his selection by stating that each teacher is free in selecting what to teach since he is the only person who is aware of his students needs. The remaining teachers offer the following ideas:

- I follow ‘All Clear’ textbook with its activities,
- I create my programme to develop specific skills such as public speaking, pronunciation, grammar, and vocabulary.
- I use different textbooks in designing my course,
- I focus on language functions and idioms.

The subsequent question was selected to confirm the previous questions, it revealed that all teachers' designed courses provided positive results and it enhanced learners' oral proficiency and competence. They provided the following justifications:

- Learners are able to express themselves in different situations,
- Learners pass oral exams successfully,
- Learners show a certain level of proficiency
- Learners

The third rubric sheds light on test design, the first question (15) revealed that teachers concentrate on accent and pronunciation training in assessing learners' oral proficiency. They added that they stress on learners' general competence and abilities, results are presented in the following graph:



Bar-Graph 3.7. Teacher's Assessment Criteria

The next question revealed that all teachers test what they taught during their designed courses. Most of them shift their classroom activities exclusively to test-type tasks such as oral interviews, role plays and describing pictures. They follow a speaking and listening test described comprehensibly by one teacher as follows:

'After giving learners' the necessary needed knowledge, the test contains two main parts speaking and listening. Speaking contains a non-stop short presentation of 3mn on one of the already introduced topics,

however, listening test contains a series of questions that learners should answer while listening.'

The subsequent question revealed that a little time is devoted to test instruction, practice and materials. All teachers believe that what they care more is the lesson with its objectives rather than testing and materials.

Besides, when asked about teachers' preferences in designing tests, all teachers declare that they prefer testing what they taught rather than using challenging questions. They justified their selection by giving the following point of views:

- If selecting challenging questions, learners will be lost and this may cause problems with the administration,
- It is very unfair to ask new questions, I prefer to test what I taught,
- Learners develop the habit of being tested on what have been done in the classroom.

Question n21 revealed that teachers wish if the administration helps them especially at material design phase. One teacher suggests if the oral exam will be changed into continuous testing throughout the whole year (*control continue*) so that learners will always be under the pressure to speak and better their oral proficiency. Another teacher posits that the oral exam should be done by a different teacher, not the one who offers the lecture to avoid any kind of subjectivity.

The last question demonstrated that testing the oral proficiency may provoke anxiety and low of confidence. The following positive and negative effects are put in the following table:

Positive Effects	Negative Effects
<ul style="list-style-type: none"> • Raising learners' oral proficiency levels, • Enhancing learners' responsibility levels, • Offering positive evaluation like praising learners' achievements • Technology-based tests 	<ul style="list-style-type: none"> • negative direct remarks from the teacher • talking in unfamiliar topics (lack of information), • tests beyond learners' level, • Anxiety and low confidence. • Receiving bad grades may lower their motivation to learn

Table 3.3. Positive and Negative Effects of Tests

3.3.5. Observation Phase I Analysis

In order to be sure of the washback impact on teaching methodology, the researcher opted for an observation of what truly happens in the classroom by inviting a peer to attend and examine what happens in the teaching of speaking. The teacher was given an observation grid to fill in and elicit useful information about the teachers' attitudes in the classroom. The following table illustrates the teacher's comment:

Statements	Teacher's Comment
Statement One	The teacher has administered a pre-teaching needs assessment questionnaire and a semi structured interview with learners to diagnose their strengths and weaknesses.
Statement Two	The course was designed to prepare students to the test. The teacher focuses on items that were part of the test.
Statement Three	The teacher explored the lesson plan very well, to the extent that students would pass the test without reviewing the lessons
Statement Four	In classroom discussion, the teacher searches only for correct answers and does not allow out of topic discussions.
Statement Five	The teacher focuses on exam content only in teaching speaking. The idiomatic expressions learned will be in the exam
Statement Six	The teacher tries to avoid challenging topics that have no relation to the speaking test; most of the time is devoted to exam-related content.
Statement Seven	The teacher does not allow further discussions about topics which are not included in the lesson plan
Statement Eight	Half of the class time is devoted to exam-related instruction.
Statement Nine	The teacher encourages dialogue memorizations,
Statement Ten	The teacher prepares her students to get all what was done in the classroom correctly in the test

Table3.4. Observation Grid Results

The teacher-observer provided the following general comments besides the grid given to him, they are summarized as follows:

- The teacher designed the course according to students needs,
- Learners' always ask about the test/exam format,
- The teacher focuses on exam-related items and always emphasizes on what the test include,
- The teacher seems to prepare learners only for scores neglecting levels.

3.3.6. Oral Proficiency Rating Analysis.

The oral proficiency scoring estimates the score of the learner based on two criteria fluency and pronunciation. Results of the students' performance on the software were entered to SPSS spreadsheet. The first test that was applied on the questionnaire was alpha Cronbach to check the internal consistency of the questionnaire. Results are displayed below:

		fluency	Pronunciation
Fluency	Pearson Correlation	1	-,024
	Sig. (2-tailed)		,896
	N	32	32
Pronunciation	Pearson Correlation	-,024	1
	Sig. (2-tailed)	,896	
	N	32	32

The results of the Pearson correlation show unacceptable results, $r = -.024$ which is a very low value. The commonly accepted values are represented in the following table:

Cronbach's alpha	Internal consistency
$0.9 \leq \alpha$	Excellent
$0.8 \leq \alpha < 0.9$	Good
$0.7 \leq \alpha < 0.8$	Acceptable
$0.6 \leq \alpha < 0.7$	Questionable
$0.5 \leq \alpha < 0.6$	Poor
$\alpha < 0.5$	Unacceptable

Table 3.5. Internal Consistency Estimations

Results show that $\alpha = -.024$, this denotes that results of this rater are not reliable. The alpha put at the beginning of the study was 5% i.e., .005. $\alpha = .896$ which means 89% of the results of this sample is rejected. Thus, automatic rater results are not valid.

In order to recheck these results' reliability, the researcher opts for another test. Results are displayed below:

One-Sample Kolmogorov-Smirnov Test

		fluency	Pronunciation
N		32	32
Normal Parameters ^{a,b}	Mean	2,9531	1,9531
	Std. Deviation	,94493	,80682
Most Extreme Differences	Absolute	,125	,196
	Positive	,122	,196
	Negative	-,125	-,148
Test Statistic		,125	,196
Asymp. Sig. (2-tailed)		,200 ^{c,d}	,003 ^c

- a. Test distribution is Normal.
- b. Calculated from data.
- c. Lilliefors Significance Correction.
- d. This is a lower bound of the true significance.

Without analyzing the sum of mean, std deviation, what catches the attention is the test significance: $\alpha = .005$ (95% the mistake percentage) is the one put at the beginning of the study, here was computed $\alpha = .200$. i.e., 200% of the results are unacceptable. Therefore, the teacher developed a personal fluency assessment grid to get to know how well students scored in their test performance in terms of proficiency. Results are displayed below:

Comprehension

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Poor	14	43,8	43,8	43,8
	Average	10	31,3	31,3	75,0
	good	8	25,0	25,0	100,0
	Total	32	100,0	100,0	

As for the first criteria, the highest percentage was in score 1 poor. Students had a poor mastery of the topic, when it comes to content that have not already been rehearsed previously. Thus, this made us think twice about the content of the oral course.

		Pronunciation			Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Poor	11	34,4	34,4	34,4
	average	12	37,5	37,5	71,9
	good	9	28,1	28,1	100,0
	Total	32	100,0	100,0	

As for pronunciation, results show the highest frequency in score 2 average. Thus, students have an average pronunciation, between a native-like and a foreign like pronunciation.

		Fluency			Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Poor	11	34,4	34,4	34,4
	Average	12	37,5	37,5	71,9
	Good	9	28,1	28,1	100,0
	Total	32	100,0	100,0	

As for fluency, results show that the highest frequency and percent was on score 2 average. Students' oral production consists of isolated words and frequently long pauses.

		Grammar			Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Poor	12	37,5	37,5	37,5
	Average	13	40,6	40,6	78,1
	Good	7	21,9	21,9	100,0
	Total	32	100,0	100,0	

Results demonstrate an average score in grammar. Thus, students show average of accuracy in using the language. They are able to handle successfully communicative tasks with average grammar mistakes.

Vocabulary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Poor	12	37,5	37,5	37,5
	average	13	40,6	40,6	78,1
	good	7	21,9	21,9	100,0
	Total	32	100,0	100,0	

Coming to the last criteria, average results were computed. Learners are able to satisfy the necessities of a basic communication learned utterances, with no use of new up-to-date vocabulary words. The mean and the standard deviation were computed and results are displayed below:

Report

	comprehension	Pronunciation	Fluency	Grammar	Vocabulary
Mean	1,8125	1,9375	1,9375	1,8438	1,8438
N	32	32	32	32	32
Std. Deviation	,82060	,80071	,80071	,76662	,76662

The mean of all the assessment criteria was close to score 2, which denotes that students have an average proficiency level. The value of discrepancy is low which denotes that the group level is probably similar, thus, students' level is homogeneous.

Besides, one sample t - Test was computed to compare means and to test whether a population mean is significantly different from some hypothesized value or not. Results are displayed below:

One-Sample Test

	Test Value = 0					
	t	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
comprehension	12,495	31	,000	1,81250	1,5166	2,1084
Pronunciation	13,688	31	,000	1,93750	1,6488	2,2262
Fluency	13,688	31	,000	1,93750	1,6488	2,2262
Grammar	13,605	31	,000	1,84375	1,5674	2,1201
Vocabulary	13,605	31	,000	1,84375	1,5674	2,1201

- *T Statistics* of this t-test is $t=13$
- *Df*: id the degree of freedom for the test. $Df=n-1= 32-1= 32$, $df=32$
- *Sig. (2-tailed)* is the p- value adapted to this study. (Sig. (2-tailed)-.000) which is perfect for this study.
- *Mean Difference* denotes the difference between the observed sample mean and the expected mean put for this study.
- *95% Confidence Interval of the Difference* is the interval for the difference between t test and the questionnaire mean.

3.4. CYCLE TWO DATA ANALYSIS

The second cycle starts two weeks after the first oral proficiency test, where the teacher has administered the same oral proficiency test two weeks after exam period to see if students have learned for the test only or not, and to compare the two results. Results are displayed below:

	N	Minimum	Maximum	Mean	Std. Deviation
HumanR	32	8,00	16,00	12,2031	1,95868
test2	32	5,00	13,00	9,7500	1,96748
Valid N (listwise)	32				

3.4.1. Washback Questionnaire Analysis

After the questionnaire distribution, each statement was coded. The scores put to each statement were entered to SPSS spreadsheets. Pre-tests were computed as shown in chapter two. The questionnaire was designed to gather data about teacher's attitudes about tests, how does the test influence their teaching. Results are put as follows:

N	Statement	Mean	St. Dev	Skewness
1	I always analyze my students' needs before I design my oral course.	4,2857	,48795	1,230
2	I develop the teaching method with which I feel more comfortable.	4,2857	,48795	1,230
3	I am satisfied with the teaching methodology I follow in teaching	3,7143	1,60357	-1,053

	speaking			
3	I follow the curriculum and the established syllabus.	3,4286	1,39728	-1,079
4	I see no importance to teach a new topic that will not be examined.	3,1429	1,21499	,414
5	My main objective is to design adequate courses that develop learners' general oral proficiency.	4,4286	1,21499	,374
6	I teach learners to prepare them be good language speakers.	4,5714	,53452	-,374
7	Examinations do not oblige me to teach selected topics.	4,4286	,53452	,374
8	I emphasize on teaching speaking rather than on scoring.	4,2857	,48795	1,230
9	I do not design my tests according to what I have taught in the classroom	2,7143	1,3801 3	,706
10	I do not test only what I taught, I keep learners expect everything to develop their proficiency.	4,1429	1,06904	-,374

Table 3.6. Teachers' Washback Questionnaire II

Skewness test is also computed to get the degree and direction of results' asymmetry, positive or negative compared to the computed means. Results are summarized below:

Statistics

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
N Valid	7	7	7	7	7	7	7	7	7	7	7
Missing	0	0	0	0	0	0	0	0	0	0	0
Mean	4,2857	4,2857	3,7143	3,4286	3,1429	4,4286	4,5714	4,4286	4,2857	2,7143	4,1429
Median	4,0000	4,0000	4,0000	4,0000	3,0000	4,0000	5,0000	4,0000	4,0000	2,0000	5,0000
Mode	4,00	4,00	5,00	4,00	2,00	4,00	5,00	4,00	4,00	2,00	5,00
Std. Deviation	,48795	,48795	1,60357	1,39728	1,21499	,53452	,53452	,53452	,48795	1,38013	1,06904
Skewness	1,230	1,230	-1,053	-1,079	,414	,374	-,374	,374	1,230	,706	-,374
Std. Error of Skewness	,794	,794	,794	,794	,794	,794	,794	,794	,794	,794	,794

The mean, is sensitive to a large value. Results displayed in the table may drive to conclude that the mean is in] 2, 72; 4, 57[comparing it to alpha set in the beginning

of the study which goes around 5, it denotes that the sum of means is close to alpha, i.e., all teachers tend to teach oral expression following the same teaching method.

As for standard deviation measures, the spread set of data in this questionnaire is large. Therefore, data of this questionnaire is most spread out. As for Skewness measures, this test shows the degree and direction of data asymmetry. The positive value of Skewness denotes a normal distribution except in statement 3, 7 and 11.

Moreover, one sample t - Test was computed to compare means and to test whether a population mean is significantly different from some hypothesized value or not. Results are displayed below:

One-Sample Test						
	Test Value = 0					
	t	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
S1	23,238	6	,000	4,28571	3,8344	4,7370
S2	23,238	6	,000	4,28571	3,8344	4,7370
S3	6,128	6	,001	3,71429	2,2312	5,1973
S4	6,492	6	,001	3,42857	2,1363	4,7208
S5	6,844	6	,000	3,14286	2,0192	4,2665
S6	21,920	6	,000	4,42857	3,9342	4,9229
S7	22,627	6	,000	4,57143	4,0771	5,0658
S8	21,920	6	,000	4,42857	3,9342	4,9229
S9	23,238	6	,000	4,28571	3,8344	4,7370
S10	5,203	6	,002	2,71429	1,4379	3,9907
S11	10,253	6	,000	4,14286	3,1542	5,1316

- *T Statistics* of this t-test is $t=4,264$
- *Df*: id the degree of freedom for the test. $Df=n-1= 7-1= 6$, $df=6$
- *Sig. (2-tailed)* is the p- value adapted to this study.
- *Mean Difference* denotes the difference between the observed sample mean and the expected mean put for this study.
- *95% Confidence Interval of the Difference* is the interval for the difference between t test and the questionnaire mean.

3.4.2. Observation Analysis Phase II

In order to be sure of the washback impact on teaching methodology, the researcher opted for an observation of what truly happens in the classroom by inviting a peer to attend and make sure of whether tests influence the ways of teaching or not. The teacher was given an observation grid to fill in and elicit useful information about the teachers' attitudes in the classroom. Teacher's comment is summarized below:

Statements	Teacher's Comment
Statement One	The teacher has an idea about learners' levels, their competences and even the scores of the first oral proficiency test.
Statement Two	The course was designed to prepare students to be able to handle any communicative situation. The teacher didn't focus on items that would be part of the test.
Statement Three	The teacher explored the lesson plan very well, and pushed learners' communicate in a very friendly atmosphere, not recalling the test.
Statement Four	The designed course help learners' become proficient language speakers.
Statement Five	In the classroom, the teacher allows learners' to discuss the proposed topic at their own will, sometimes the discussion becomes a debate where all students are engaged, not calling any test item.
Statement Six	The teacher does not teach any test item, the focus is on engaging learners' to speak.
Statement Seven	The teacher proposes thought provoking questions to push learners' get in the track of the discussion
Statement Eight	All the class time is devoted to participation in the classroom discussion.
Statement Nine	The teacher does not encourage dialogue memorizations at all.
Statement Ten	The teacher prepares her students to speak about any topic that had not been already rehearsed.

Table 3.7. Observation II Grid Results.

Besides, the observer-teacher adds the following noted important remarks

- The teacher introduces daily new topics to students, asking them to bring each session a current event from newspapers and speak about it for 10mn.
- The teacher also introduces an updated technique to boost learners' motivation, confidence and wellbeing and incorporate *acts of kindness* in the classroom.

- This technique helps them develop their social integration and motivation to speak in English about their good deeds.

3.4.3. Oral Proficiency Test Results

Assessment was a one-to-one interview between the teacher and the student. Scoring was based on four assessment stages as explained below:

Stage	Description	Scoring	Criteria
I	Warm-up	2,5	See Learner's Readiness
II	Level-check	10	Cpmprehension 2 Content 2 Vocabulary2 Fluency 2 Pronunciation2
III	Probes	5	unexpected challenging questions
IV	Wind- Down	2,5	Conclusions of the interview.

Table 3.8. Assessment Grid

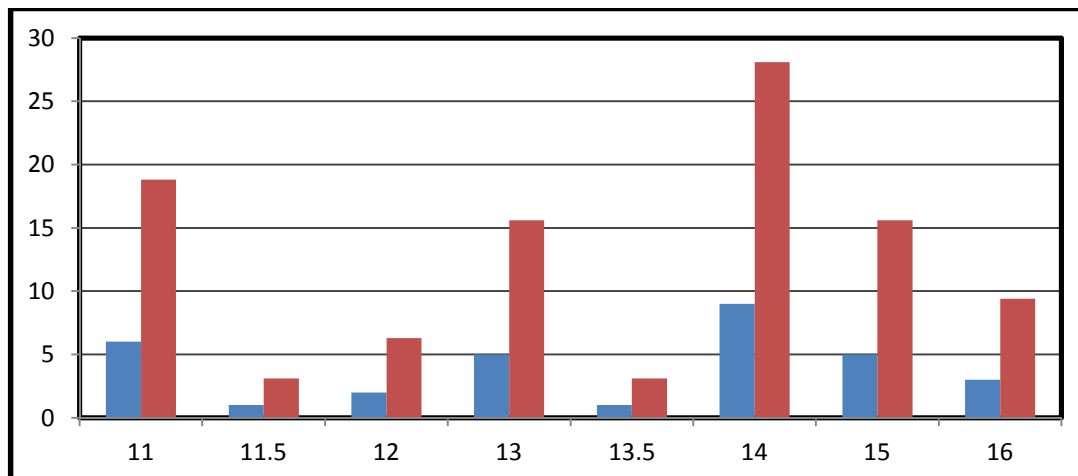
Learners' scores varied form 11/20 to 16/20. The mean (the measure of central tendency) and the standard deviation (the measure of the dispersive tendency) were calculated and the following table demonstrates the students' scores:

Descriptive Statistics							
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
PROFICIENCY2	32	11,00	16,00	13,4063	1,59858	-,185	,414
Valid N (listwise)	32						

The lowest score was Min=11 and the highest score was 16. The overall mean of students was 13.40 which denote an average central tendency score. As for the Standard Deviation, the dispersive tendency, it was 1.595 which denotes a large amount of variation in oral proficiency scores. Subsequently, the researcher analyzed the frequencies of scores in the oral proficiency test, results are displayed below:

PROFICIENCY2					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	11,00	6	18,8	18,8	18,8
	11,50	1	3,1	3,1	21,9
	12,00	2	6,3	6,3	28,1
	13,00	5	15,6	15,6	43,8
	13,50	1	3,1	3,1	46,9
	14,00	9	28,1	28,1	75,0
	15,00	5	15,6	15,6	90,6
	16,00	3	9,4	9,4	100,0
Total		32	100,0	100,0	

As it is displayed in the table above, the highest percentage 28,1 % was in score 14, which denotes an average score of oral proficiency. Results of the human rater scores are put below:



Bar-Graph 3.8. Test Scores

3.4.4. Proficiency Rating Scale

As in the first oral proficiency rating scale, the scoring estimates the score of the learner based on two criteria fluency and pronunciation. Since results of the automatic rater showed a low reliability, the teacher opted for a human scoring only. Results of the students' performance on the software were entered to SPSS

spreadsheet. The teacher has developed a personal fluency assessment grid to get to know how well students scored in their test performance in terms of proficiency.

As for the first criteria, the sum of 46.9% of students had a good mastery of the topic when it comes to content that have not already been rehearsed previously.

comprehension

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Poor	1	3,1	3,1	3,1
	average	12	37,5	37,5	40,6
	Good	15	46,9	46,9	87,5
	excellent	4	12,5	12,5	100,0
	Total	32	100,0	100,0	

As for pronunciation, results show the highest frequency in score 14 good score. Thus, students have good pronunciation, between a native-like and a foreign like pronunciation.

Pronunciation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Poor	6	18,8	18,8	18,8
	Average	12	37,5	37,5	56,3
	Good	14	43,8	43,8	100,0
	Total	32	100,0	100,0	

As for the fluency, results show that the highest frequency and percent was on score 16 average score. Students' oral production consists of clear acceptable utterances.

Fluency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Poor	5	15,6	15,6	15,6
	Average	16	50,0	50,0	65,6
	Good	11	34,4	34,4	100,0
	Total	32	100,0	100,0	

Results demonstrate a good score in grammar. Thus, students show a average of accuracy in using the language. They are able to handle successfully communicative tasks with average grammar mistakes.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Poor	3	9,4	9,4	9,4
	average	8	25,0	25,0	34,4
	Good	14	43,8	43,8	78,1
	excellent	7	21,9	21,9	100,0
	Total	32	100,0	100,0	

Coming to the last criteria, good results were computed. Learners are able to satisfy the necessities of a basic communication learned utterances, with no use of new up-to-date vocabulary words.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Poor	2	6,3	6,3	6,3
	average	5	15,6	15,6	21,9
	Good	17	53,1	53,1	75,0
	excellent	8	25,0	25,0	100,0
	Total	32	100,0	100,0	

The mean and the standard deviation were computed results are displayed below:

	N	Minimum	Maximum	Mean	Std. Deviation
Comprehension	32	1,00	4,00	2,6875	,73780
Pronunciation	32	1,00	3,00	2,2500	,76200
Fluency	32	1,00	3,00	2,1875	,69270
Grammar	32	1,00	4,00	2,7813	,90641
Vocabulary	32	1,00	4,00	2,9688	,82244
Valid N (listwise)	32				

The mean of all the assessment criteria was more than score 2.5 which denote that students proved to have a good proficiency level. The value of discrepancy is low

which denotes that the group level is probably similar. Students' level is positively homogeneous. Besides, One sample t - Test was computed to compare means and to test whether the mean is significantly different from some hypothesized value or not. Results are displayed below:

	Test Value = 0					
	T	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
comprehension	20,605	31	,000	2,68750	2,4215	2,9535
Pronunciation	16,703	31	,000	2,25000	1,9753	2,5247
Fluency	17,864	31	,000	2,18750	1,9378	2,4372
Grammar	17,358	31	,000	2,78125	2,4545	3,1080
Vocabulary	20,419	31	,000	2,96875	2,6722	3,2653

- *T Statistics* of this t-test is $t=17$
- *Df*: id the degree of freedom for the test. $Df=n-1= 32-1= 31$, $df=31$
- *Sig. (2-tailed)* is the p- value adapted to this study. (*Sig. (2-tailed)*-.000) which is perfect for this study.
- *Mean Difference* denotes the difference between the observed sample mean and the expected mean put for this study.
- *95% Confidence Interval of the Difference* is the interval for the difference between t test and the questionnaire mean.

3.4.5. Post-Test Questionnaire Analysis

The questionnaire was designed to gather data about learners' attitudes about tests, how they prepare for the test, does the test influence their oral proficiency. The mean the standard deviation and Skewness values are displayed in the table below:

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	1	3,0	3,1	3,1
	Disagree	1	3,0	3,1	6,3
	Agree	16	48,5	50,0	56,3

	Strongly Agree	14	42,4	43,8	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

As for the first criteria, the highest percentage was in score 3 agree. Students all agreed and noticed that the course was designed to help them get ready for the test. Results demonstrate that students all strongly agree that the exams influence their learning process. 37.5 % believe that exams influence their learning. Exams force them to study certain topics. Results are put as follows:

S2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	2	6,1	6,3	6,3
	Disagree	5	15,2	15,6	21,9
	Strongly Disagree	1	3,0	3,1	25,0
	Agree	12	36,4	37,5	62,5
	Strongly Agree	12	36,4	37,5	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results demonstrate that 42.4 % of students all strongly agree that the current oral course help me improve my oral proficiency

S3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	3	9,1	9,4	9,4
	Disagree	4	12,1	12,5	21,9
	Agree	11	33,3	34,4	56,3
	Strongly Agree	14	42,4	43,8	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results demonstrate that 63,6 % of students all strongly agree that learning test-taking techniques is most important in class.

S4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	11	33,3	34,4	34,4
	Strongly Agree	21	63,6	65,6	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

As for pronunciation, results show the highest frequency in score 5. Thus, students have a belief that performance on the test reflects their oral abilities.

S5

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	10	30,3	31,3	31,3
	Disagree	3	9,1	9,4	40,6
	Agree	6	18,2	18,8	59,4
	Strongly Agree	13	39,4	40,6	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

As for the test, results show that the highest frequency and percent was on score 4. Students agree that Examinations influence their learning process every year.

S6

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	3	9,1	9,4	9,4
	Disagree	6	18,2	18,8	28,1
	Strongly Disagree	6	18,2	18,8	46,9
	Agree	10	30,3	31,3	78,1
	Strongly Agree	7	21,2	21,9	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results demonstrate that students all strongly agree that the exams influence their learning process. 30.3 % got worried immediately before taking the test.

S7

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	5	15,2	15,6	15,6
	Disagree	7	21,2	21,9	37,5
	Strongly Disagree	1	3,0	3,1	40,6
	Agree	9	27,3	28,1	68,8
	Strongly Agree	10	30,3	31,3	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results demonstrate that students all strongly agree that the exams influence their learning process, 39.4 believe that they focus on their performance rather than scores.

S8

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	3	9,1	9,4	9,4
	Disagree	4	12,1	12,5	21,9
	Agree	12	36,4	37,5	59,4
	Strongly Agree	13	39,4	40,6	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results demonstrate that 30.3% of students all strongly agree that they prefer having my oral test using the computer to record their answers.

S9

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	3	9,1	9,4	9,4
	Disagree	6	18,2	18,8	28,1
	Strongly Disagree	10	30,3	31,3	59,4
	Agree	8	24,2	25,0	84,4
	Strongly Agree	5	15,2	15,6	100,0
	Total	32	97,0	100,0	

Missing	System	1	3,0	
Total		33	100,0	

Results demonstrate that 33, 3% of students all agree that they start feeling uneasy just before getting the test score.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	5	15,2	15,6	15,6
	Disagree	6	18,2	18,8	34,4
	Strongly Disagree	3	9,1	9,4	43,8
	Agree	11	33,3	34,4	78,1
	Strongly Agree	7	21,2	21,9	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

As for the test, results show that half of students felt uneasy upset feeling while others did not have such a feeling.

S11

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	6	18,2	18,8	18,8
	Disagree	10	30,3	31,3	50,0
	Agree	10	30,3	31,3	81,3
	Strongly Agree	6	18,2	18,8	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results of statement12 show that 30. 3% students seem to defeat themselves while working on tests like speaking test.

S12

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	8	24,2	25,0	25,0
	Disagree	5	15,2	15,6	40,6
	Strongly Disagree	1	3,0	3,1	43,8

	Agree	10	30,3	31,3	75,0
	Strongly Agree	8	24,2	25,0	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results of the thirtieth statement show that 30.3% of students seem to defeat themselves while working on tests like speaking test.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	5	15,2	15,6	15,6
	Disagree	8	24,2	25,0	40,6
	Strongly Disagree	9	27,3	28,1	68,8
	Agree	5	15,2	15,6	84,4
	Strongly Agree	5	15,2	15,6	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results show that 27% of students thought about the consequences of failing on the test and 27% didn't think of the consequences of failing.

S14

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	5	15,2	15,6	15,6
	Disagree	9	27,3	28,1	43,8
	Strongly Disagree	4	12,1	12,5	56,3
	Agree	9	27,3	28,1	84,4
	Strongly Agree	5	15,2	15,6	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results show that that 36. 4% of students got so nervous that they forgot facts they knew.

S15

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	3	9,1	9,4	9,4
	Disagree	6	18,2	18,8	28,1
	Strongly Disagree	1	3,0	3,1	31,3
	Agree	10	30,3	31,3	62,5
	Strongly Agree	12	36,4	37,5	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results show that 27.3% of students agreed that the more they prepare for the test the more confused they get.

S16

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	5	15,2	15,6	15,6
	Disagree	9	27,3	28,1	43,8
	Strongly Disagree	4	12,1	12,5	56,3
	Agree	9	27,3	28,1	84,4
	Strongly Agree	5	15,2	15,6	100,0
Total		32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results show that 39, 4% of students got perplexed that they forgot facts they knew, to defeat themselves while working on tests like speaking test.

S17

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	3	9,1	9,4	9,4
	Disagree	7	21,2	21,9	31,3

	Strongly Disagree	3	9,1	9,4	40,6
	Agree	6	18,2	18,8	59,4
	Strongly Agree	13	39,4	40,6	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results show that 54, 5% of students agree that they are able to handle successfully different communicative skills.

S18

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	3	9,1	9,4	9,4
	Strongly Disagree	1	3,0	3,1	12,5
	Agree	18	54,5	56,3	68,8
	Strongly Agree	10	30,3	31,3	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

As for the subsequent statement, Results show that 39. 4% of students prefer one-to one interview with the teacher rather than an online web based assessment.

S19

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	6	18,2	18,8	18,8
	Disagree	4	12,1	12,5	31,3
	Strongly Disagree	1	3,0	3,1	34,4
	Agree	8	24,2	25,0	59,4
	Strongly Agree	13	39,4	40,6	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Accordingly, results show that 30.35% of students disagree about having some thoughts of doing poorly that interfered with their concentration on the test.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	9	27,3	28,1	28,1
	Disagree	10	30,3	31,3	59,4
	Strongly Disagree	4	12,1	12,5	71,9
	Agree	5	15,2	15,6	87,5
	Strongly Agree	4	12,1	12,5	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Therefore, results show that 24,2% of students are not sure about their feelings after the test.

S21

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	8	24,2	25,0	25,0
	Disagree	6	18,2	18,8	43,8
	Strongly Disagree	6	18,2	18,8	62,5
	Agree	5	15,2	15,6	78,1
	Strongly Agree	7	21,2	21,9	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results of statement22 show that 30. 3% students show that they are unsure while taking the oral course.

S22		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	5	15,2	15,6	15,6
	Disagree	8	24,2	25,0	40,6
	Strongly Disagree	5	15,2	15,6	56,3
	Agree	10	30,3	31,3	87,5
	Strongly Agree	4	12,1	12,5	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results of statement23 show that 21% of students believe that the speaking test does not bother them so much.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	6	18,2	18,8	18,8
	Disagree	7	21,2	21,9	40,6
	Strongly Disagree	7	21,2	21,9	62,5
	Agree	6	18,2	18,8	81,3
	Strongly Agree	6	18,2	18,8	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Results show that 39,4% of students prefer talking tests rather than face o face interview with the teacher.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Sure	5	15,2	15,6	15,6
	Disagree	4	12,1	12,5	28,1
	Strongly Disagree	3	9,1	9,4	37,5
	Agree	7	21,2	21,9	59,4
	Strongly Agree	13	39,4	40,6	100,0
	Total	32	97,0	100,0	
Missing	System	1	3,0		
Total		33	100,0		

Besides, Skewness test is computed to get the degree and direction of results' asymmetry, positive or negative compared to the computed mean. Moreover, one sample t - Test was computed to compare means and to test whether a population mean is significantly different from some hypothesized value or not. Results are displayed below:

One-Sample Test						
Test Value = 0						
	T	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
S2	27,260	31	,000	4,28125	3,9609	4,6016
S3	17,084	31	,000	3,84375	3,3849	4,3026
S4	16,336	31	,000	3,90625	3,4186	4,3939
S5	54,583	31	,000	4,65625	4,4823	4,8302
S6	10,415	31	,000	3,28125	2,6387	3,9238
S7	14,812	31	,000	3,37500	2,9103	3,8397
S8	12,571	31	,000	3,37500	2,8275	3,9225
S9	16,383	31	,000	3,87500	3,3926	4,3574
S10	14,987	31	,000	3,18750	2,7537	3,6213
S11	13,076	31	,000	3,28125	2,7694	3,7931
S12	11,458	31	,000	3,00000	2,4660	3,5340
S13	11,240	31	,000	3,15625	2,5836	3,7289
S14	12,607	31	,000	2,90625	2,4361	3,3764
S15	12,407	31	,000	3,00000	2,5068	3,4932
S16	14,886	31	,000	3,68750	3,1823	4,1927
S17	12,407	31	,000	3,00000	2,5068	3,4932
S18	13,962	31	,000	3,59375	3,0688	4,1187
S19	20,437	31	,000	4,00000	3,6008	4,3992
S20	12,715	31	,000	3,56250	2,9911	4,1339
S21	10,295	31	,000	2,53125	2,0298	3,0327
S22	10,885	31	,000	2,90625	2,3617	3,4508
S23	12,858	31	,000	3,00000	2,5242	3,4758
S24	11,975	31	,000	2,96875	2,4631	3,4744
S25	13,366	31	,000	3,59375	3,0454	4,1421

- *T Statistics* of this t-test is $t=10,290$
- *Df*: id the degree of freedom for the test. $Df=n-1= 7-1= 6$, $df=6$
- *Sig. (2-tailed)* is the p- value adapted to this study.
- *Mean Difference* denotes the difference between the observed sample mean and the expected mean put for this study.
- *95% Confidence Interval of the Difference* is the interval for the difference between t test and the questionnaire mean.

3.5. INTERPRETATIONS OF THE MAIN RESULTS

Taking into account the results attained in the previous section, this part will be deduced to the summary and interpretation of the results obtained vis-à-vis the research questions raised at the onset of this investigation. This research involves representing and analyzing different research variables, where the relationship include a dependent variable *learners oral proficiency*, and an independent variable which is the teaching methodology followed by the teacher which is adopted depending on the test washback, a positive and a negative washback.

The researcher at this level will mostly rely on statistical logic to interpret the results achieved. This is mainly due to the fact that statistical interpretation will result in logical, rational and balanced interpretations and help draw satisfactory conclusions. Thus, regression analysis, ANOVA, correlation significance analysis, and indirect proportion will be the most adopted statistical methods used in interpreting the results.

Clearly defined, regression analysis is a quantitative method of testing the relationships between the dependent variable (oral proficiency) and the independent variable (the teaching method). The regression model adopted in this study specifies basically the relation between the dependent variable (Y) to a combined function of the independent variable (X) and some unknown influencing factors (α). The equation is:

$$Y \approx f(X \alpha)$$

The regression equation is used to predict:

- Values of Y: if both X_1 , X_2 are known variables thus, the possible values of Y can be calculated.
- To know the most influencing factor of the variables X_1 and X_2

The formula of the regression equation is as follows:

$$Y = \alpha + \beta X_1 + \varepsilon$$

$$Y = \alpha + \beta X_2 + \varepsilon$$

It should be noted that there must be enough data to estimate a regression model. Thus, the estimated variables were put into the following formula:

$$\text{Oral Proficiency} = \alpha + \beta \text{Washback}_1 + \beta \text{washback}_2$$

The first research question revolved around how positive/negative washback influence learners' oral proficiency? It intended to explore the influence tests exerts on learners' oral proficiency. Thus, in order to examine statistically this regression, simple linear regression equation was formulated as follows:

$$Y = \alpha + \beta X_1 + \varepsilon$$

$$Y = \alpha + \beta X_2 + \varepsilon$$

Results are displayed below:

Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,732 ^a	,186	,024	1,24336

a. Predictors: (Constant), sum

This model summary provided the R and R² values. The R=.432 which presents a high degree of correlation between learner's oral proficiency and the teaching method followed by the teacher. R² value presents how much the total variations in the independent values are.

The subsequent table represents ANOVA, which reports how well the regression equation fits into the presented data, results are shown below:

ANOVA^a

Model	Sum of Squares	Df	Mean Square	F	Sig.
1 Regression	1,770	1	1,770	1,145	,003 ^b
Residual	7,730	5	1,546		
Total	9,500	6			

a. Dependent Variable: PROFICIENCY2

b. Predictors: (Constant), washback

This table indicates that the regression equation is significantly well designed. This is clear from the regression model run $0.003 < 0.005$, which indicates that the regression model is statistically significant. Coming to the coefficient table, it provides data to predict whether there is a statistical contribution to the model or not. By looking at the sig values. Results are displayed below:

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
1 (Constant)	15,220	1,674		9,090	,000
washback	-,051	,047	-,432	-1,070	,003

a. Dependent Variable: PROFICIENCY2

The regression equation is put as follows:

$$Y = 15,22 + -0.43X_1$$

From the results displayed above, findings denote that there is a statistical association between the dependent and the independent variables, the teaching method followed by the teacher and learners' oral proficiency. The more the test has a negative washback on teaching the more the oral proficiency decreases. The more the test has a positive washback on teaching, the more proficient speakers they become.

Hence, when teachers follow an ordinary teaching not exam- related instruction, student' oral proficiency increases, and their speaking achievement increases and vice versa. This confirms the first hypothesis.

The second question was about seeking possible correlations existing between the teaching method followed by the teacher and the test scores. Here, in order to examine statistically this correlation, since the data was ordinal, the researcher opts for *Pearson* correlation coefficient to determine the correlations between variables. Results are displayed below:

		SCORES	WASH
SCORES	Pearson Correlation	1	,789
	Sig. (2-tailed)		,000
	N	32	7
WASH	Pearson Correlation	,859	1
	Sig. (2-tailed)	,002	
	N	7	7

Thus, $r = .856$. To statistically determine whether this value represents a positive relationship between the two variables, it should be compared with the *correlation value standards* (See Appendix G). A positive value of r means a positive correlation which stands for a significant correlation between the variables. Conversely, a negative value of r means a negative correlation which denotes a negative accordance between the variables. The following table clearly elucidates this:

$r \leq -1$	$r \geq +1$
Negative Correlation	Positive correlation
No significant relationships between variables	Strong relationship between variables

Table3.9.Correlation Significance

Therefore, our findings denote that there is a strong positive statistical relationship between test washback and learners' test scores. Thus, as the test has a high impact on teaching, the scores are influenced too. Therefore, there is a direct proportion between the two variables, and this strongly confirms Hypothesis II.

Coming to the third question, which revolved around the possible relations existing between learners' test scores and their proficiency levels, it intended to explore whether good scores denote good oral proficiency or not. Thus, in order to

examine statistically this regression, simple linear regression equation was formulated as follows:

$$Y = \alpha + \beta X_2 + \varepsilon$$

Results are displayed below :

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-12,092	2,674		4,523	-,053
	PROFICIENCY2	,058	,198	,054	,295	-,072

a. Dependent Variable: sum

Table3.77. B Constant Coefficient

The regression equation is put as follows:

$$Y = -12,09 + -0.54X_2$$

From the results displayed above, findings denote that there is no statistical association between the dependent and the independent variables, oral proficiency and the test scores. Sig value is negative which means that there is no influence between the two variables. Results from the previous section denote that scores do not reflect their oral proficiency. And this rejects the third hypothesis.

Trying to link the three variables together throughout this research work, it was found that there is a strong correlation relationship of the three variables oral proficiency, washback. Results reveal that there is a positive relationship between washback and the oral proficiency.

Oddly enough, there is no significant relationship between learners' oral proficiency and their test scores. Thus we may conclude that test scores do not give accurate image about the oral proficiency of learners. Thus, there is a direct proportion between the three variables, and this is displayed below:

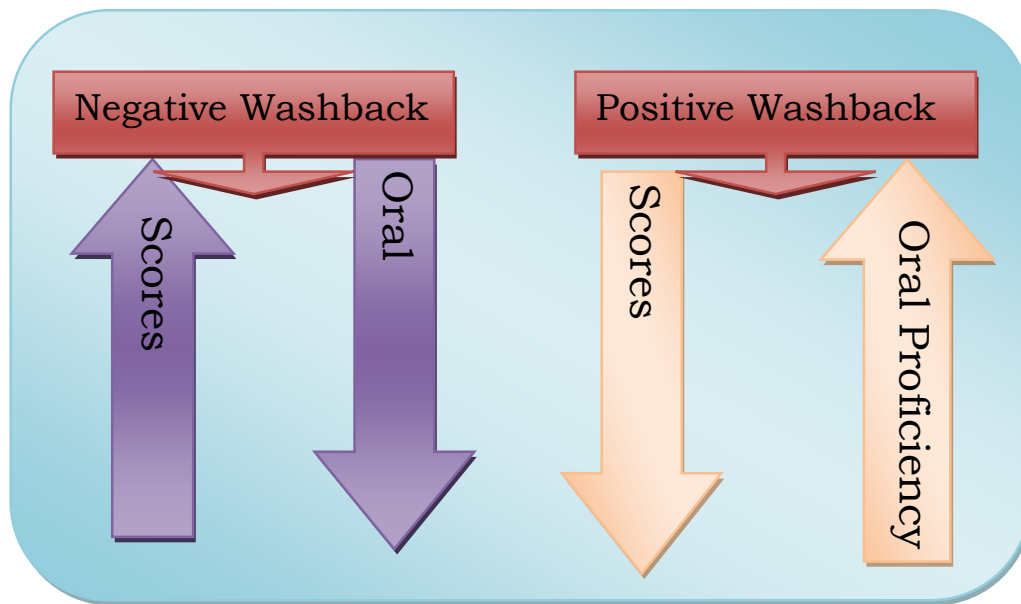


Diagram3.1 Indirect Proportion between Variables.

3.6. CONCLUSION

This chapter presented data analysis and interpretation part of this research work, it tried to answer the research questions using a variety of research instruments. However, findings denote that there is no statistical association between the dependent and the independent variables, oral proficiency and the test scores.

Therefore, teachers are asked to focus more on learners' needs and expectations from the course rather than teaching them exam-related topics. Teachers should cultivate ordinary teaching to avoid negative washback so that learners will not only have good scores in the test only but also a good language proficiency that will allow them become good language speakers in previously unrehearsed communication contexts. Hence, the subsequent chapter offers practical considerations and suggestions that may support the teacher and the learner within the classroom setting.

**Chapter Four
Considerations and
Implications for Innovation**

4.1. INTRODUCTION.....	140
4.2. NEW PARADIGMS FOR TESTING: <i>Call for Urgent Progress</i>	140
4.3. IMPLICATIONS FOR CHANGE: <i>Teachers as Professional Assessors</i>	143
4.4. LAB-BASED ASSESSMENT: <i>Technology Introduction</i>	149
4.4.1. Testing Speaking Reconsidered.....	151
4.4.2. Online Assessment in Oral Testing: Fine Grading Techniques.....	152
4.5. Introducing Online Blended Teaching In Time of Covid-19 Pandemic	161
4.6. FACTORS INFLUENCING LANGUAGE ASSESSMENT.....	167
4.6.1. Test Anxiety as a Barrier in Testing.....	168
4.6.2. Factors Affecting Test Anxiety.....	171
4.7. NEW DIRECTIONS TO WASHBACK PHENOMENON.....	174
4.8. CONCLUSION.....	175

4.1. INTRODUCTION

Language teachers routinely encounter within their daily experience pervasive human emotions that interfere with having a suitable learning environment. This chapter intends to help both teachers and learners enjoy a relaxing testing environment where teachers are intellectually involved in the classroom so that learners get better achievements. Hence, based on the results achieved in the previous chapter, the following part attempts to shed light on a number of implications that are worth discussing. Viewing the necessity to change the way teachers teach/assess speaking skills, the researcher maps out a suitable land to innovate build an enjoyable relaxing classroom environment where both students and teachers are actively involved. With this idea in mind, it attempts to summarize some prefatory comments needed before starting a speaking course. This chapter tries to map out the terrain for constructing a speaking-based lab course for EFL learners, including pronunciation practice, and taking into account some psychological variables within the teaching/learning process.

4.2. NEW PARADIGMS FOR TESTING: *Call for Urgent Progress*

Testing is gaining great attention more and more and thus awareness must be driven on new language testing frames. In an important insertion, Chomsky argues about the threat of standardized testing on students' performances and achievements. He states that *the assessment itself is completely artificial and that:*

The people sitting in the offices, the bureaucrats designing this, they're not evil people, but they're working within a system of ideology and doctrines that turns what they're doing into something extremely harmful.

Chomsky (2017:23)

Thus, designing standardized tests stops creativity and interest and even imagination; *“the student can't pursue things, maybe some kid is interested in something, can't do it because you got to memorize something for this test tomorrow. And the teacher's future depends on it, as well as the student”* (Chomsky, 2017). Thus, performance in such exams is not an indication of anything. He also claims that people cannot be judged by some *“artificial standards”*. In this vein, Diane Marie states that: *“One of the most distressing characteristics of education reformers is that*

they are hyper-focused on how students perform, but they ignore how students learn.” Thus, these anxiety providing tests do not measure real goals of education and therefore, as put by Ayers

Standardized tests can't measure initiative, creativity, imagination, conceptual thinking, curiosity, effort, irony, judgment, commitment, nuance, good will, ethical reflection, or a host of other valuable dispositions and attributes. What they can measure and count are isolated skills, specific facts and functions, the least interesting and least significant aspects of learning. Ayers, (1993: 116):

Therefore, a radical rethinking assessment policy in Algeria is necessary and the section of the *forward-thinking* that must take place now in the testing arena is to realize that anxiety providing High-stakes exams are not the sole miracle recipe of evaluating schools and students. In this sense, as Nicholes and Berliners point out:

Finland, the highest achieving country in the world in reading, mathematics and science, have [sic] no standardized tests that resemble ours whatsoever, though they use teacher made tests in their classroom and school accountability system. Their system uses high standards for allowing teachers into the profession, awards high pay and bestows high status to those that enter teaching, provides rigorous and extensive professional development for the teachers, and depends on trusting relationships to improve academic achievement. Clearly there are highly successful models of how to build a national school system that we should study before assuming that our corrupting high-stakes accountability system is the only one that will work.

Nicholes and Berliners (2005: 165-166)

To avoid that, educators must opt for new testing techniques where the learners learn from the test and not only become slaves to it.

The implementation of new testing paradigms has been the concern of language assessment researchers over the last few decades. However, real-life application of concepts like ‘*autonomy*’ ‘*collaboration*’ and ‘*student centered learning*’ is

impractical within a high-stakes context. Due to this impracticality problem, standardized tests around the world continue to use summative, product-oriented tests to local educational systems which have been found to be pedagogically lacking behind.

Notwithstanding the importance of such exams on students' prospective lives, the teaching act, here, becomes more *test-driven*, motivation becomes extrinsic and teachers design their courses with a great focus on *test taking* strategies, and *exam English*.

Engaging students in assessment is not a new notion in TEFL professionals. Oddly enough, the traditional view of assessment is roughly teacher-centered i.e., the teacher is the external examiner who design and administers the test. However, within a new State-of-the-art procedure, one may opt for principles of autonomy students'-centeredness to assessment where the student is involved in assessment at every level. This leads to the concept of "*learning conversations*" (Harri-Augstein & Thomas, 1991), in which students discuss the learning that has occurred up to that point, and make further learning goals based upon their conclusions. In this fashion, Andrew (2005:17) alternative assessment tools may be cited as follows:

- *Portfolios Assessment* offers an approving means of looking at students' speaking processes (Hauck, 1994). With portfolio assessment, learners save the recordings they have done over the term, including early trials. They then compare the two performances to know the progress they have made. The following step is to arrange the selected recordings to compile a collection that will make up the portfolio in which their findings and progress is presented.

- *Self /Peer Assessment*, this type of assessment replaces teacher assessment. Self and peer feedback is a common practice nowadays, by fellow advice, learners better understand and grasp criteria reach success

- *journals and learners' diaries* can be successfully employed in language assessment

- *Process-Based Assessment or naturalistic assessment* where teaching is integrated within assessment. The teacher sets up a non-threatening environment, with portfolios, journals, self-assessment (Finch,2005)
- *Online assessment or Web-Based Assessment*, for instance, PLATO EduTest Assessment is a web-based assessment program used to store results and diagnose students' strengths and weaknesses, learners' needs identified by objectives and standard. Feedback is immediate and the link can be found here <http://www.edutest.com>

4.3. IMPLICATIONS FOR CHANGE: *Teachers as Professional Assessors*

Language teaching involves a number of complicated paradigms that contribute to the variability and intricacy of teaching. Such complicated image leads to a challenging role that teachers should act wisely to get good results. These roles move from a course designer and developer to test designer. Thus, foreign language teachers always face challenging situations where he is supposed to perform taking different dimensions into consideration.

A glance at the nature of tests, they are described as instruments designed by people for people in a specific distinctive socio-cultural setting; therefore, in order to understand how a test might influence learning, teachers must become professional test designers to be able to consider the target audience and the system to effectively contribute to the construction and use of the test. The question to be raised at this level, how can teachers reach such a challenging target? And what measures should be taken into account to design suitable tests?

Looking at the matter from a specialist view point, it is found that language assessment researchers such as McNamara (2001) and Edelenbos and Kubanek-German (2004) believe that as language teachers/testers, one should broaden the scope of research to encompass classroom assessment and become professional language assessors, this is to meet the needs of teachers and learners and better the testing mission. In fact, how teachers diagnose the strengths and weaknesses of their learners would also contribute to a better understanding of what can or could be diagnosed.

Perhaps one of the most important study on teachers as competent testers is that by Edelenbos and Kubanek-German (2004:260), who comprehensively develop the notion of a *teacher's diagnostic competence*. They discussed how teachers can develop 'the ability to interpret foreign language growth in individual children'. It is to be noted that language teachers should become familiar with the new methods of assessment and testing and concentrate on the fact that learners' competence is changeable and it differs through time.

From a more contextualized fashion, and considering the fact that testing creates major challenging task to teachers, it is strongly agreed that assessing speaking will add extra focus from teachers' part as put by Ludenberg (1929: 195): '*oral skills are less measurable because they are less tangible, more subject to variation, and probably will involve the cumbersome and time-consuming expedient of the individual oral examination*'. From the results attained within this research work, it was found that speaking teachers often relate learners' oral proficiency to scoring. However, Bachman and Palmer 1996; Winke, Gass and Myford 2013, state that '*test-takers' oral production could at least be redressed through clearly defining the focal construct and explicitly addressing sources of construct-irrelevant variance in rater training (i.e. factors that should have no bearing on raters' scoring decisions*'

Considering the above statements, and based on the collected literature, it would be wiser to state that in order for teachers to become professional testers and test designers, they have to broaden their knowledge and areas of research on assessment challenging tasks. For instance, the British Council (2018) puts forward a number of criteria to assess speaking professionally, it is stated that:

If you need to assess your students' speaking skills there are a number of factors you need to consider. These include considering the test taker and their needs, the kind of information you want to know about their speaking skills and the most appropriate ways to elicit and assess these skills so that your test is appropriate and accurately scored.

Inspired by these factors, teachers should ask a number of questions before, during and after the test.

- *What is involved in the speaking skill?*
- *What are the issues raised when assessing speaking?*
- *What are we assessing?*
- *What are the different dimensions we need to consider about speaking?*
- *What aspects of speaking must any assessment attempt to replicate?*
- *How are we assessing?*
- *What are the different exam formats we can use to assess speaking?*
- *What are some of the different tasks we can use to assess speaking?*
- *What are the different settings we can use to assess speaking?*
- *How do examiners award scores?*
- *How can technology improvements help test design?*
- *How can you make sure a speaking task for them is fair?*

Answers to these questions will construct a more comprehensive picture to test design/score/assessment. In a clearer picture, the following ideas were suggested as answers to the asked questions by the British Council (2018):

When assessing professionally, teachers should bear in mind that speaking involves:

- *Getting a message across,*
- *Organising a message for the listener,*
- *Using a range of language for different purposes,*
- *Being accurate so there are no misunderstandings,*
- *Being fluent so a listener is not 'waiting',*
- *Being intelligible,*
- *Contributing to keeping the conversation going.*

When assessing, the raised issues might be:

- *Being objective,*
- *Giving reliable assessment across different test takers,*
- *Timing,*
- *Making a task clear and unambiguous.*

As Suggested Test Formats, Interviews, Oral Presentations, Interactive Tasks and Group Discussion are practical tasks for learners. These tasks might be performed in the following ways:

- *Performed live in front of one or more examiners,*
- *Recorded and evaluated later by one or more examiners,*
- *Talk over the phone to an examiner or to a recording device,*
- *Communicate via a computer or tablet, with the examiner either live or working from a recording.*

✓ *Extemporaneous speaking*

This kind of activity promotes participating using the target language in, it is a speech carefully prepared, but delivered with no notes or text. The speaker has to use improvisational conversations, in which he should adapt his speech according to the audiences' attention. This kind of speech is tricky, the learner has to be well prepared no to forget words and phrases. The speaker has to involve some note cards only to prompt and guide his speech from one point to another. No memorisation is allowed just notes to know which idea to tackle next.

✓ *Oral Presentation or Lecture*

Delivering an oral presentation is crucial in developing the *habit to speak*, according to Emden and Beker (2004: 1), "*developing the abilities to speak to an audience is one of the greatest benefits you'll ever get from your time in further or higher education.*" thus, oral presentations in front of an audience help learners' improve their oral proficiency levels. Here are myriad of oral presentation. By understanding the aim of the presentation one can decide on the type to employ.

✓ *Informative Oral Presentation*

This type of presentation is intended to deliver an informative presentation; it can be either through teaching content or presenting a product. The speaker has to present ideas in sequence in a very limited time. Chivers and Shoolbred (2007:.5) say

that “*This type of presentation [informative presentation] is used in many organizations where students are expected to report progress at key stages of a project.*” All in all, this type of presentation should inform the audience and answer their questions accordingly.

✓ *Group Discussion:*

Group discussion is the process of talking between two or more groups of students to achieve specific objectives about a certain topic. Ur (1997:2) says that “*thinking out some problems and situations together through verbal interchange of ideas is simply called as to discuss*”. This discussion is an opportunity to exchange information, opinion, and ideas. It has many positive sides, Ornstein (2000) states that dividing students into small group provides an opportunity to enhance students’ cooperation and social skills develop their critical thinking techniques, become more actively engaged in learning as for as teachers are concerned they are to better monitor students’ strengths and weaknesses.

• *Role play (and real play)*

Role play is an academic activity and one of the communicative techniques used to teach learn and assess second language learning. It provides students with many opportunities to perform a defined role with the purpose of acquiring competence in the learned language. According to Larsen-Freeman (2008:112) “*role-plays are very important in the communicative approach because they give students an opportunity to practice communicating in different social contexts and in different social roles*” Songco (2002) and Siddiqui (2008) affirm that there are five steps for teachers to follow in order to organize and implement Role play in oral classroom so as to enhance students’ oral communication production. The table below summarises this:

Stages	Content
The Briefing stage	Devide roles between members of the group Selection of the situation
The Checklist Stage	Deal with how students will perform each scene of the play
The Interaction Stage	This stage gives an opportunity to students to perform the play and rehearse, it should contain critical observation of the performed scenes.
The Forum Stage	Creates a discussion between members of the play. The aim is to solve communication problems and the gaps aiming to find solutions to the depicted problems.
The Debriefing Stage	The most important stage in role play. It concerns the assessment stage where remarks, reactions and comments are addressed about the performance.

Table 4.1. Different stages of Role play

Using role plays in promoting oral interaction has many advantages on learners' oral proficiency. Ladousse, (1987:7) on his side claims that: *“Role play enables students not just to acquire set phrases, but to learn how interaction might take place in a variety of situations”* .this means that role play can be implied to develop their learning as well as their social skills. The figure below summarizes this:



Figure 4.1. Advantages of Role Play Adapted From Chesler And Fox (1966)

- *Oral Portfolios*

Reviewing the benefits of alternative assessment techniques, oral assessment portfolios present prospective advantages to the learners. This alternative assessment tool promotes autonomy, self reflection and self evaluation. An oral portfolio assesses students speaking performances at intervals and involves students in the assessment process. Students are asked to put all their recordings videos in their portfolios as being their raw materials. Similarly, portfolios enable learners to handle their learning. When learners are required to collect what they have learned and presented they start looking at learning from a very different angle. Correspondingly, Johnson and Rose (1997:76) stated that: “*Without self-assessment, portfolios become merely another storage area for student work*” Johnson and Rose (1997:76). Thus, portfolios push students to acquire new skills among which self analysis and self reflection.

4.4. LAB-BASED ASSESSMET: *Technology Introduction*

In the light of intensive modern research and with the advent of technologies in the world in general, and education in particular, the urgent demand for *professional instruction* has been improved noticeably, making it enormously challenging for English teachers and educators to meet and prepare their learners grow and develop within this new scenario. A growing subject of research and hot debates in the field of technological education are witnessed within this changing time of globalisation. An increased attention continued to be prominent as far as speaking assessment and online assessment are concerned.

Information and Communication Technologies (ICTs) continues to grow in importance in developing the curriculum of higher education. Technology-enhanced classrooms seem to enable multi-modal teaching, change curricula and spawn rich forms of online research and collaboration; it also may promote discovery learning, learner autonomy, and learner-centeredness and create motivation in the classroom. Bringing technology into our classrooms is, in fact, the concern of a number of researchers who call for a pause for reflection to provide opportunities to consider the possible impacts of new technologies on the learning outcomes.

Tremendous developments in technology have brought out new needs for human beings to facilitate and speed up the process to progress. In view of this fact, foreign language teachers have always been ahead of the curve in integrating technology in their teaching and assessment. Our main concern in the present section is to demonstrate how technology-based instruction may enable students to learn at their own speed, give and receive feedback from peers and instructors alike.

Our argument is that the implications of this new global order have created novel patterns of institutions, creating new structures, new learning opportunities as well as new issues within teaching-learning processes. In fact, incorporating technology in assessment in general and in assessing speaking in particular will create a great challenge for teachers. As a first step for innovation, they should change their teaching visions by upgrading the curriculum. Given the realities of globalization, knowledge work, and accelerating societal change, it seems obvious that our learners are flooded with devices and online options: *what* they learn as well as *how* and *when* they learn is changing. Thus, our teaching programme should respond to these changes by redesigning the curriculum adequately through upgrading the curriculum. In this vein, before inserting technology in assessment, teachers should first teach using technology.

Our main aim, this is to revisit speaking assessment and redefine our course design objectives, and the way we assess our students. Multi-dimensional goals should be set regarding the teaching curriculum and the assessment strategy. Teachers should treat the learning environment as being an active process, redefine knowledge, promote high level of critical thinking, and emphasise both on the how and the what, the process and the product. there has been a recent shift in pedagogy to alternative methods of assessment, which, among other things, is believed to enhance learners' metacognitive knowledge and strategies leading to the development of lifelong learning skills (Council of Europe, 2001).

In the last few decades, there has been a recent shift in educational institutions from traditional methods of language teaching to alternative methods. Thus, speaking about reforms and innovation in teaching without speaking about assessment seems to be outdated. In a more contextualised manner, technology introduction in language

teaching and assessment is believed to enhance students metacognitive knowledge and strategies. Beyond the fact that students feel more engaged and motivated using technological devices, (I-pads, tablets, smart phones...), teachers may create *energetic and engaging lessons*. Every lesson must be enticing, energetic, and fun. It is easy for students in any class to be distracted and lose interest in such context, having opportunities to disturb and create discipline issues with their mates. This is especially true in a large classroom where the teacher will be unable to control and manage a huge number of learners, thus, as a short-term solution, technological aids in lessons may be unique, and full of attention grabbers.

However, nothing seems to be perfect, machine scoring systems are not likely to completely supplant assessments of L2 speech. In fact, it is ultimately humans who will remain forever the perfect evaluator who are at the level of judging all communication components and the intended message. Thus, human judgments are likely to stay the standard against any automated speaking assessment system. Besides, *'as technology continues to revolutionize the nature of human communication and to open up new interactional possibilities on a global scale (Kramsch 2012), the need to perform and assess complex speaking tasks in reliable and valid ways will continue to persist'*.

4.4.1. Testing Speaking Reconsidered

A glance through the current assessment strategies would give us an interesting picture of the varied strategies followed by teachers in assessing speaking. Learners' oral competence is generally assessed by the extraction of speech from learners in a face-to-face speaking test or computer-assisted oral tests which represent a new trend that has entered the field of language assessment and didactics. With the advent of the ELT industry and the move towards a communicative language teaching scenario, there was an urgent needs for a large body of research about oral proficiency assessment and communicative measures (Turner, 1998). Over the last few decades, there have been a growing number of researches publications and studies consisting our stockpile of knowledge about new trends in oral language assessment.

Viewing oral proficiency testing arena, researches and ELT publications have been focusing mainly on developing a set of assessment criteria oriented towards communicative competence model of oral proficiency, interactive language proficiency eliciting teachers to have face to-face oral interaction. Viewing the matter from a more contextualized manner, schools and universities are attempting to integrate the communicative competence model into language speaking assessment. There are several assessment tasks proposed by teachers for assessing their learners' oral proficiency (Chinda, 2009) proposed that teachers may have an idea about his learners' proficiency through their responses to oral questions slips, reacting to visual prompts, storytelling, presenting an oral presentation, role play, speaking in a talking circle. Ultimately, direct performance based tasks, face-to-face interview, role play are commonly the most popular choices among teachers for assessing speaking. As far as role plays are included, both scripted and non-scripted ones, learners are given written roles/prompts and are asked to present the scenario, they are given time to prepare and rehearse before acting it in the class in front of their friends. On the other side, non-scripted ones are delivered to students to perform the activity based on a non-rehearsed preparation, they are given some time to prepare and rehearse the written script.

4.4.2. Online Assessment of Oral Proficiency: Fine-Grading Techniques

Online assessment as a new trend in language testing is gaining ground more and more and thus awareness must be driven on new language testing frames. New innovative techniques of language assessment have arisen with the technological boom that we are witnessing. Web-based assessment has been extensively applied in the field of language education. It is gaining ground more and more due to the increasing demand for innovation in the field of language teaching and research (chang:2013). Electronic examinations are beneficial because of the easiness of administering the tests, easiness of scoring function offering immediate correction and instant information regarding learners' scores and easiness of grading. Advantages of using online assessment display a great significance to teachers and educational researchers, they give timely feedback, lightening teachers' teaching

loads, facilitating e-Learning and fostering students' ability of self-assessment (Wang: 2011).

With the advent of artificial intelligence, the ability to store and process a large amount of information has caused facilities to language teachers around the world. Among those facilities lay down speech analysis techniques in grading. Through the use of computer-based softwares, computers now can process and analyze learners' speech similar to humans. Many smart phones can receive commands in the form of a complex set of algorithms and statistical data. conversely, web-based assessment software is not only used to recognize someone's exists level *per se*, but also to teach them how to speak In a more contextualized manner, one of the outstanding softwares used in this research is *speechace software*, although used with its free trial mode, it proved to be efficient. The following figure represents the website frame:



Figure 4.2. Speechace Platform

This platform proposes a set of offers mainly about learners' fluency and pronunciation assessment, the speechace guidelines claim that “*Our patented technology is unique in its ability to score a learner's speech and pinpoint individual syllable and phoneme level mistakes in a user's pronunciation in real time*” (adopted from speechace website).

Chapter Four: Considerations and Implications for Innovation

As for assessment the website presents the best in-class speech technology experience designed with care to score pronunciation and fluency. Under the following layer:

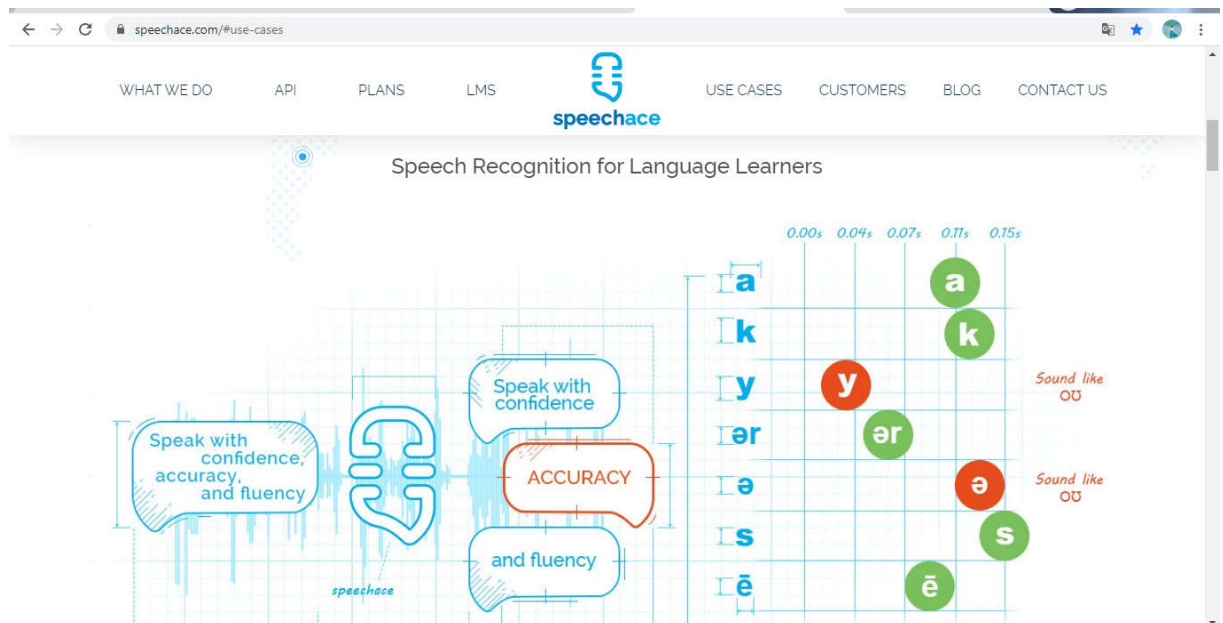
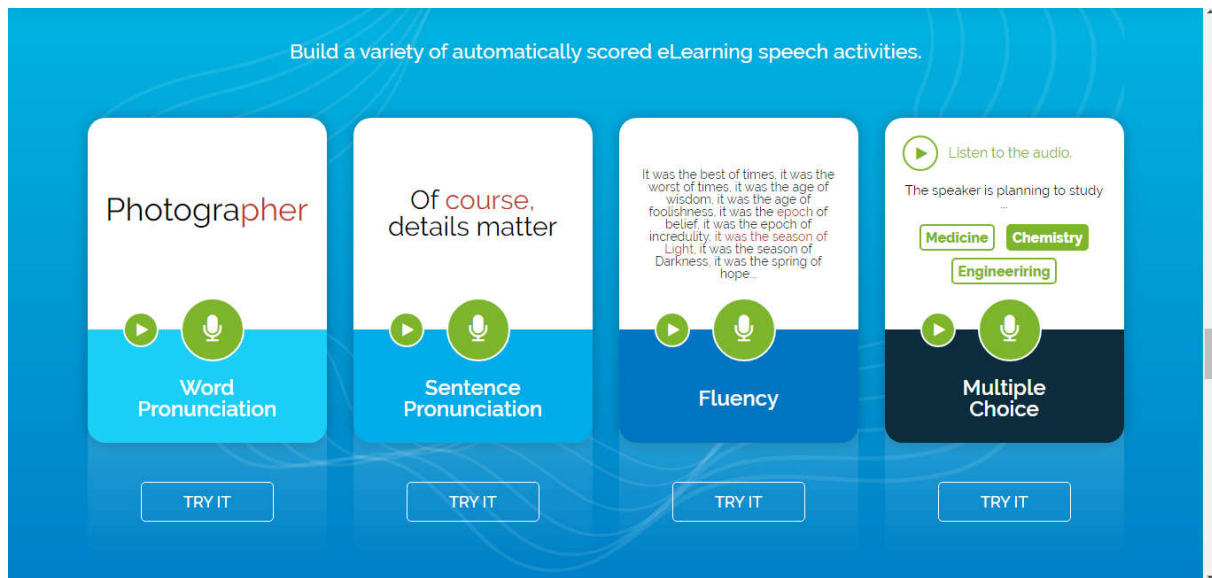


Figure 4.3. Speech Recognition Page

This application offers a variety of of automatically scored e-learning activities among which the one presented below:

BASIC	PRO	PREMIUM
Ideal for short word and sentence pronunciation activities	Ideal for speaking fluency, IELTS and PTE	Ideal for longer speaking activities and custom needs
GET STARTED	GET STARTED	GET STARTED
Scoring <ul style="list-style-type: none"> Pronunciation 	Scoring <ul style="list-style-type: none"> Pronunciation Fluency IELTS & PTE 	Scoring <ul style="list-style-type: none"> Pronunciation Fluency IELTS & PTE
Feedback <ul style="list-style-type: none"> Sentence Word Syllable Phoneme 	Feedback <ul style="list-style-type: none"> Sentence Word Syllable Phoneme Speaking rate Pausing Length of Run IELTS Speaking estimate PTE Speaking estimate 	Feedback <ul style="list-style-type: none"> Sentence Word Syllable Phoneme Speaking rate Pausing Length of Run IELTS Speaking estimate PTE Speaking estimate
Languages <ul style="list-style-type: none"> US English UK English 	Languages <ul style="list-style-type: none"> US English UK English 	Languages <ul style="list-style-type: none"> US English UK English

Figure 4.4. Speechace Pronunciation Offers



Figures 4.5. Speechace Speech Activities

The first speech activity develops learners’ word pronunciation, where learners are asked to choose words of a given topic like presented below:

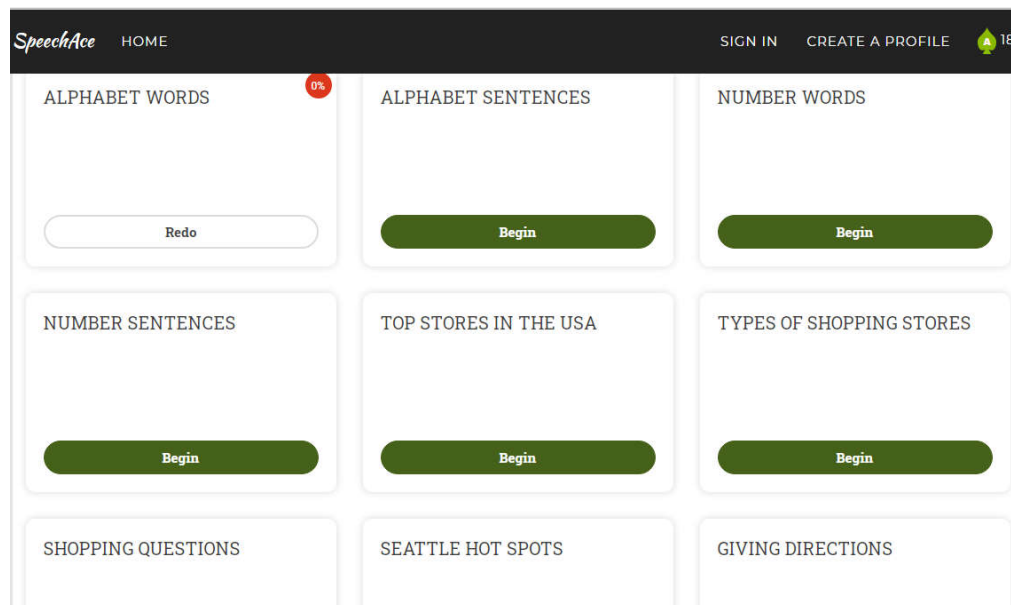


Figure 4.6. Speechace Word Pronunciation Topics

The second option is about sentence pronunciation. The student is asked to read carefully the sentence provided, record it and then get the right pronunciation of words, the right intonation and then scores. Example is provided below:

▶ Yes, I am. I found this apartment online. Is there a bank near here?



Figure 4.7. Speechace Sentence Pronunciation


As for the most important part the fluency activity. Speechace offers pronunciation training through fluency correction. Students are asked to read the provided text, with respect to pronunciation rhythm intonation word stress, and punctuation.

2 of 5 Quit

▶ Press play to listen to the question and then record your answer

Well, one reason is that when you travel alone, you are free to do exactly what you want. You don't have to consider another person's interests and you can plan your itinerary as you wish. Another reason is that it is easier to make friends when you are travelling by yourself.

People are more likely to talk to you when you are on your own. So, travelling on your own can help you make new friends.




Skip

3 of 5 Quit

▶ Press play to listen to the question and then record your answer

Yes, I do. Travel today is vastly different than what it used to be. In the past, a traveller had little idea about what to expect when they arrived at their destination. These days, the internet connects our world in ways previous generations could only dream about.

We can instantly review destination information and make travel arrangements. Also, in the past, people could only travel by land or sea and travelling was often long and unsafe. Today, international flights provide a quick, safe and affordable travel experience to any destination in the world.



Skip

Figure 4.8. Speechace Fluency Assessment

As an example from one student's task is the following figure. It shows where student failed to pronounce words, missed the intonation. Each mistake is written in red so the student might click and listen to the native speakers' pronunciation, and then re record himself until he/she gets it right. A score is finally given to the student, on following the IELTS Rating Scale.

The figure consists of two screenshots of a digital speech fluency worksheet interface. Each screenshot shows a listening exercise with a question and a text passage containing red text indicating pronunciation errors. The top screenshot is labeled '2 of 5' and shows a score of 6.5 on a pink background. The bottom screenshot is labeled '3 of 5' and shows a score of 7.0 on a green background. Both screens include a 'Continue' button and a microphone icon for recording.

2 of 5 Sounds **30** Quit

▶ Press play to listen to the question and then record your answer
Tap any sentence to improve. Be sure to practice and improve the RED ones.

▶ Well, one reason is that when you travel alone, you are free to do exactly what you want.⁶³ You don't have to consider another person's interests and you can plan your itinerary as you wish.⁶⁴ Another reason is that it is easier to make friends when you are travelling by yourself.⁷⁵

People are more likely to talk to you when you are on your own.⁷⁷ So, travelling on your own can help you make new friends.⁶⁸

6.5 IELTS score

Continue

3 of 5 Sounds **44** Quit

▶ Press play to listen to the question and then record your answer
Tap any sentence to improve. Be sure to practice and improve the RED ones.

▶ Yes, I do.⁶² Travel today is vastly different than what it used to be.⁶⁵ In the past, a traveller had little idea about what to expect when they arrived at their destination.⁶⁹ These days, the internet connects our world in ways previous generations could only dream about.⁷²

We can instantly review destination information and make travel arrangements.⁶⁷ Also, in the past, people could only travel by land or sea and travelling was often long and unsafe.⁶⁶ Today, international flights provide a quick, safe and affordable travel experience to any destination in the world!⁶¹

7.0 IELTS score

Continue

Figure 4.9. Speechace Fluency Worksheets

The application helps in assessing many criteria of the same recording. Learners have first to login and record their audios speaking about a certain given topic. The speechace app helps assessing a set of criteria. Among which:

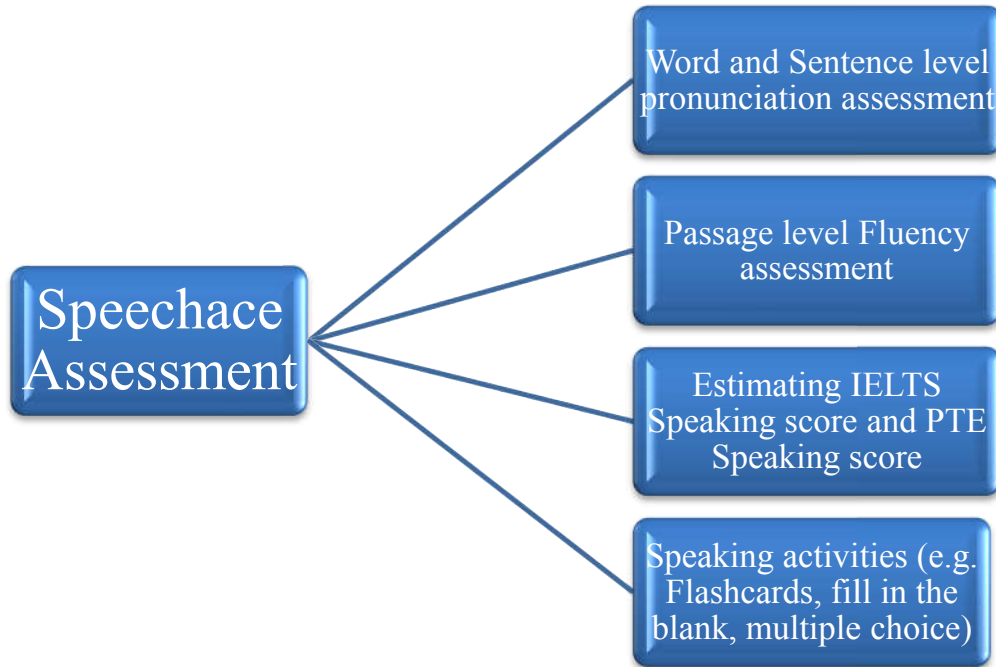


Diagram 4.2. Speechace Assessment Criteria (adopted from speechace.com)

Speechace assessment provides a quality score for pronunciation of the recorded speech, syllable and phoneme. This would give an interesting facility in scoring and feedback about pronunciation mistakes. Assessment fields are summarized below in the adopted table:

Field	Description
Quality Score	An overall pronunciation score for the the entire utterance on a scale of 0 to 100..
Syllable Score List	a list of syllables in each word in the word score list, each with its own quality score
Word Score List	a list of words in the utterance, each with its own quality score

Syllable List	Score_	a list of syllables in each word in the word_score_list each with its own quality score
Phone List	Score	a list of phonemes in each word in the word score list[], each with its own quality score
Extent		start and end boundaries of a syllable or phoneme in units of 10 msec.

Table 4.2. Fluency Assessment Criteria adopted from Speechace.com

The Speechace measures the accuracy of learners' pronunciation of a sentence, word, phoneme, and syllables on a rating scale from 0 to one hundred. The following table summarizes the quality scores of how Speechace functions to provide accurate results about learners' performance. This is put in the following table:

Score	Color	Description
90 - 100	Green	Excellent. Native or native-like
80 – 90	Green	Very Good and clearly intelligible.
70 – 80	Orange	Good. Intelligible but with one or two evident mistakes.
60 – 70	Red	Fair. Possibly not intelligible with several evident mistakes.
0 – 60	Red	Poor and must be reattempted.

Table4.3. Scoring Criteria (adopted from Speechace.com)

For a better understanding, the speechace fluency API has a number of assessment criteria included. They are presented in the following diagrams:

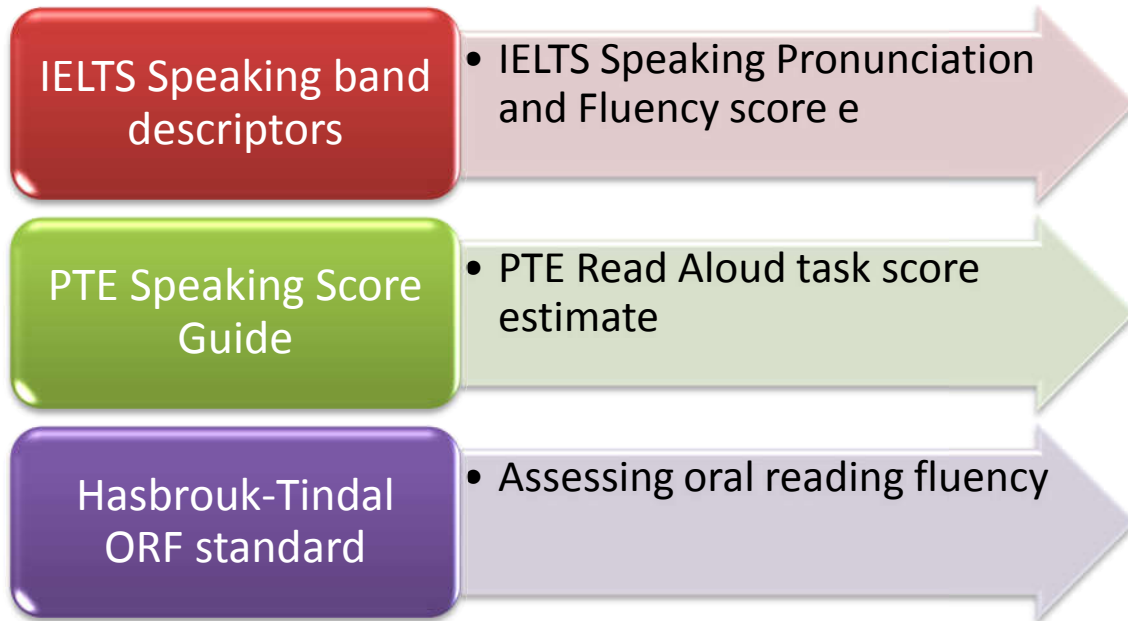


Figure 4.10. API Assessment Guide (retrieved from speechace.com)

Taking to account the above score estimations, the API provide segment by segment scores facilitating getting easy feedback to help the teacher score students recordings by practicing rehearsing and repeating. An example of this assessment activity based on the IELTS speaking practice is shown below:

2 of 5 Words Quit

▶ Press play to listen to the question and then record the answer below
Tap any sentence to improve. Be sure to practice and improve the RED ones.

▶ Not really, but I would love to be one.^{8.5} I am the youngest in the family, so growing up, I got ordered around by my two brothers all the time.^{8.5} Now I am older, and I want to voice my opinion, but they don't always take my opinion seriously.^{8.5} Being the baby in the family, I was well protected, loved, and bossed around by them.^{8.5} I heard the birth order myths.^{7.5} I need to figure out a way for my brothers to take me seriously.^{7.5} Right now, I am not succeeding.^{8.5}

8.0 IELTS score 🎤 Continue

Figure 4.11. Students IELTS Scoring

✓ *Videoconferencing as a Tool for Developing Oral Proficiency*

Speaking is a daunting activity for many foreign language learners' mainly those who have a poor literacy and a poor stockpile of knowledge. Conversely, learners are born in a digital world; they are natives of digital literacy. Ultimately, Innovation continues in language teaching research aided by dynamic advancements in technology. Thus, teachers must meet learners where they are. Video conferencing is a mode of developing learner's oral proficiency which is taking ground more and more. Since it allows people from different parts of the world to bridge the gap and communicate. It is now used for teaching purposes to help learners' overcome the problem of limited opportunities for speaking practice in the classroom by offering the opportunity to communicate with native speakers.

4.5. Introducing Online Blended Teaching in Time of Covid-19 Pandemic

With the crisis deteriorating, the number of published articles is increasingly getting high. Herein, in a very recent interview on the first of April 2020 with the American linguist Noam Chomsky, who was self-isolated in Louisiana USA, he reported that:

“COVID-19 is a very glaring new illustration of the cruelty neoliberal capitalism. The SARS epidemic was overcome 15 years ago.....The viruses were identified, sequenced, vaccines were available. Labs around the world could have started working right then on developing protection for potential coronavirus pandemics. But they didn't. Why did they not do it? Because the market's signals were wrong..... We have handed over our fate to drug companies – private tyrannies, corporations, which are unaccountable to the public.”

Thus, this pandemic had been expected long before its advent, but the cruel imperatives of an economic system in which there's “no profit in preventing a future catastrophe” prevented efforts to plan for such a crisis. Oddly enough, he claims

that companies would rather produce body creams because they are much more profitable, instead of finding a vaccine that will protect people from a total destruction. He asserts that labs could have found a vaccine for the pandemic by now, but *the neoliberal plague has blacked that*. Ironically, he notes the recent reports that Cuba, who suffered from western sanctions for long time ago, is sending doctors to help European countries to fight the pandemic. He said:

I mean this is so shocking that you don't know how to describe it. Germany can't help Greece but Cuba can help European countries. Just as when you see thousands of people dying in the Mediterranean, fleeing from countries Europe has devastated for centuries and being sent to their deaths in the Mediterranean, you don't know what words to use!

www.express.co.uk (03/04/2020)

The covid-19 pandemic has started in Algeria by March 2020 and has soon developed into a serious situation. In response to this situation, with the rapid spread of this pandemic, a number of decisions have been made at a national scale. As for the educational sector, Chitour, the Minister of Higher education urged to lockdown universities, and asked to take all measures to stop the spread of the virus, he states that:

Algeria is living, like almost all the countries of the world, an exceptional situation marked by a real coronavirus pandemic (covid19), which encourages all institutions (states, institutions and private individuals) to take all possible measures to effectively and responsibly prevent the risks of the spread of this epidemic.

As the Covid-19 pandemic continues to affect our daily lives, teachers around the world have found themselves running virtual classes, interacting with their students through social media channels, and learning by doing as they provide distance education to over 1.5 billion students impacted by the Covid19 pandemic's confinement. There exist now a plethora of reliable, well-designed and

Chapter Four: Considerations and Implications for Innovation

valid digital assessment tools that may be used to facilitate communication, learning and assessment in the time of crisis. Nevertheless, our stakeholders are experiencing today opportunities for a paradigm shift in the ways teachers teach, and the way learners learn.

Promoting Continuity of Learning using technological devices for teaching and for assessment is a highly required aspect within times of pandemics. The digital devices coupled with innovative assessment instruments like e-portfolios, videoconferencing, Google classrooms and other teaching/ testing platforms are offering genuinely versatile and highly robust ways to get in touch with students in the exceptional crises, ensuring continuity of education for learners at all levels. Furthermore, shifting the emphasis from physical delivery methods to online virtual methodologies, learners are offered “*assessment when ready*” alternatives. This provides teachers with more efficient control over learners’ assessment.

Converting to online teaching materials in very limited period of time has been a challenge for teachers during this period of pandemic. However, it is to be noted that the learning itself does not change; what changes is the ways the delivery of the knowledge. Thus, teachers need to understand the shift from classroom based instruction into online based instruction. The following table summarizes the shift:

Face-to-face classroom	Online classroom
Lectures	<ul style="list-style-type: none">• Pre-recorded presentations with screen recording software and/or webcam• Live sessions using video software like Zoom, WebEx, CollaborateVideo, etc.• Web pages, shared documents, & other media
Learning resources & handouts	<ul style="list-style-type: none">• YouTube, Vimeo, Khan Academy, & other third-party links• Files uploaded and shared via LMS, email, or shared drive
Teacher-to-student communication	<ul style="list-style-type: none">• Email or chat• Instructor announcements• Discussion, assignment, and quiz feedback
Student-to-student communication	<ul style="list-style-type: none">• Email or chat• Online discussion boards• Discussions using chat or video software

Group work	<ul style="list-style-type: none"> • Offline group projects using collaborative documents • Online discussion boards • Group work using chat or video software
Office hours	<ul style="list-style-type: none"> • Open office hours using chat or video software • One-on-one student meetings using chat or video software
Assignments & assessments	<ul style="list-style-type: none"> • Assignment submissions via email or LMS • Online asynchronous discussions • LMS or document-based quizzes • Graded and non-graded online activities
Student presentations	<ul style="list-style-type: none"> • Live presentation using video software • Recorded presentation submissions
Scheduling	<ul style="list-style-type: none"> • Shared calendar applications • Collaborative sign-up document
Face-to-face classroom	Online classroom
Lectures	<ul style="list-style-type: none"> • Pre-recorded presentations with screen recording software and/or webcam • Live sessions using video software like Zoom, WebEx, CollaborateVideo, etc. • Web pages, shared documents, & other media
Learning resources & handouts	<ul style="list-style-type: none"> • YouTube, Vimeo, Khan Academy, & other third-party links • Files uploaded and shared via LMS, email, or shared drive
Teacher-to-student communication	<ul style="list-style-type: none"> • Email or chat • Instructor announcements • Discussion, assignment, and quiz feedback
Student-to-student communication	<ul style="list-style-type: none"> • Email or chat • Online discussion boards • Discussions using chat or video software
Group work	<ul style="list-style-type: none"> • Offline group projects using collaborative documents • Online discussion boards • Group work using chat or video software
Office hours	<ul style="list-style-type: none"> • Open office hours using chat or video software • One-on-one student meetings using chat or video software
Assignments & assessments	<ul style="list-style-type: none"> • Assignment submissions via email or LMS • Online asynchronous discussions • LMS or document-based quizzes

	<ul style="list-style-type: none"> • Graded and non-graded online activities
Student presentations	<ul style="list-style-type: none"> • Live presentation using video software • Recorded presentation submissions
Scheduling	<ul style="list-style-type: none"> • Shared calendar applications • Collaborative sign-up document

Table 4.4. Online Teaching Vs Face-to-Face Classrooms (Adopted from pearson.com)

Moreover, among the best well designed platforms is Pearson platform. It offers content, assessment and digital services for teachers and learners. This website helps provide learners with necessary skills. It provides an accurate discrimination of the difference between online teaching and ordinary classroom instruction. The following figure shows the platform:

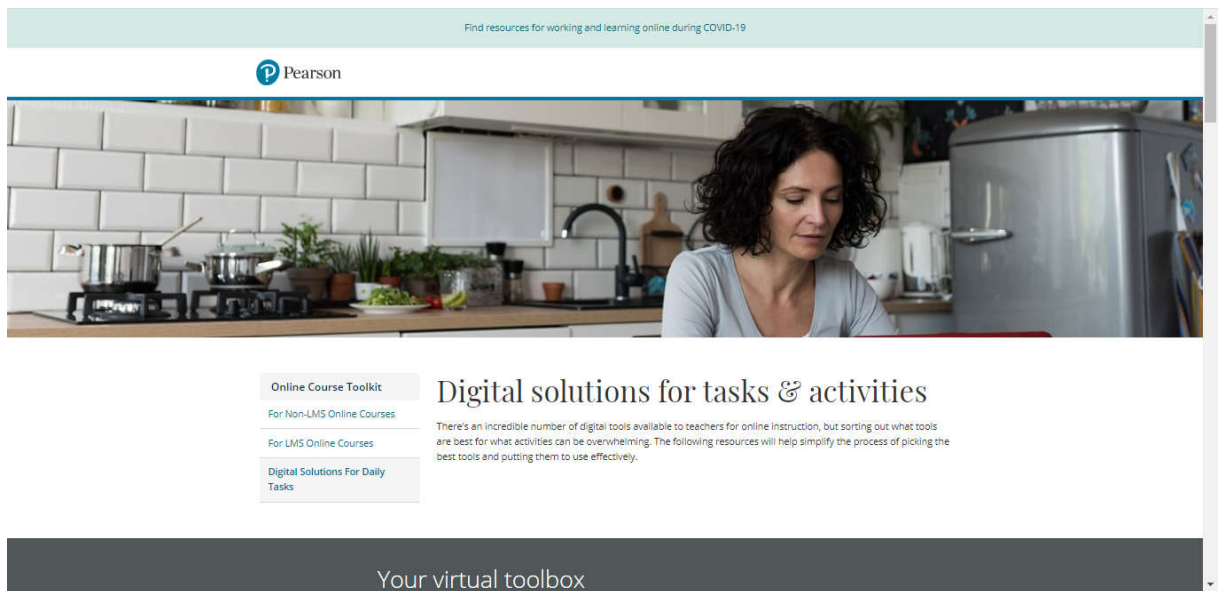


Figure 4.12. Pearson.com

At the time when the Ministry of Higher Education and Scientific Research is trying hard to save the university season with the advent of covid-19 pandemic, which caused the compulsory closing of higher education institutions and stopping the normal life of citizens, by activating distance education in which all the universities of the country were involved. They launched e-learning platforms; however, the debate now is on the lack of infrastructures and capabilities for students especially those in remote areas. Some student organizations, including the Free Student Union, see that the accreditation of teaching via digital platforms has met with great dissatisfaction

Chapter Four: Considerations and Implications for Innovation

among students due to the great delay in digitizing the sector and the majority of universities are lacking the necessary tools of distance education.

Additionally, the former Minister of Higher Education and Scientific Research, Tayeb Bouzid, gave a piece of advice to students to overcome this ordeal, saying: *“You and your parents, in these difficult circumstances, may be very anxious, due to this unique and unprecedented situation that we all live in. There is no doubt that you ask yourself many questions regarding your studies, and about the best ways to save your college year.”* He continued, *“I know that a large part of you does not have the means that allow permit you to follow online lessons, real and interactive lessons.”*

It is now acknowledged that it is impossible at the time of the pandemic to ensure ordinary teaching to students. The former minister of higher education suggested a number of pedagogical operations to be taken into account to enable students, at home, obtain their lessons to ensure their continuity of learning. To this end, an urgent adoption of coherent strategy that includes four important aspects needs to take place. In his Facebook page he posted the following points:

Aspects	Actions
Psychological Aspect	It is necessary that teachers, although confined at home, sign their recovery reports posted online by their respective institutions. This can give teachers and administration the sensation that the work is actually done, and can also allow administrations to watch and supervise the work provided by teachers. Teachers must also make sure that online lessons should be easily accessed; students must also sign an online attendance sheet. Telecommunication Companies must offer internet packages at competitive prices for students and even for teachers.
Educational Aspect	Teachers must permit students to recreate or reorganize the courses previously sent in Word or PDF format through a series of practical exercises. The student must have the possibility to get into direct contact with his teacher (professional email) in order to consult him. Psychologists must support students and teachers who have lost

	parents during the pandemic or who themselves have been suffering. They must also accompany students suffering from stress due to their fear of missing their academic year because of their confinement or their inability to take online courses.
Health Aspect	Doctors and nurses, in collaboration with the wilaya health services, must set up a campus health plan (universities, university campuses, student housing, etc.) for a possible safe return for students, and keep the first responsible i.e. the Rector informed of all actions taken.

Table 4.5. Strategies to deal with learning in time of Pandemic (adopted from formal minister of higher education page)

4.6. FACTORS INFLUENCING LANGUAGE ASSESSMENT

Notwithstanding the importance of testing in language teaching, it has become all-encompassing in current educational culture. While testing may seem to offer an excellent way to accurately assess students’ progress, it is believed that testing is affected by numerous factors that impact not only the students who take the tests, but also parents, teachers, and schools. Therefore, the quest to find the factors influencing teaching within testing can be a daunting task, and certain factors may be the result of the teaching strategy followed by the teacher in his classroom. In a more intricate manner, researchers agree on the fact that testing appears to be anxiety-provoking, and teachers need to consider the psychological side of learners when being tested..

Given the diversity of factors affecting learners’ psychology, Oxford (1996:98) assumes that “*The affective dimension of learning is probably the most significant variables which may influence the language learning success or failure*”. Thus, these affective variables play a crucial role in the process of language attainment, positive emotions and attitudes may turn the teaching more enjoyable and effective conversely, negative emotions can inhibit the learning process and its development.

Despite the proliferation of literature in the field of educational psychology, little seem to be done so far in the way anxiety affects language learning and testing. According to Aida (1994:165), research on foreign language anxiety still is underdeveloped and:

...studies examining the relationship between anxiety and learner characteristics will help us increase our understanding of language learning from the learner's perspective and provide a wider range of insights.

Herein, the application of psychology in language teaching had been the interest of many researchers all around the world. In view of this, a great number of researchers (Bailey, 1966; Horwitz, 2005; Scarason, 1996) seem to be aware of the urgent need to search about the impact of these psychological variables on the learning process. Among these psychological factors are the affective factors with “*anxiety*” as one of their sub-categories. Anxiety was found to be a stunt in language performance in some studies, anxiety was shown to be facilitative to language learning.

4.6.1. Test Anxiety as a Barrier in Testing

Language Anxiety has been a matter of considerable interest in language education setting for educators since it is a major obstacle to foreign language learning that the learners need to overcome (Cheng, 2008).

Research has exposed that anxiety is not infrequent in almost all disciplines of learning. Recently, Cassady (2010:1) brings up the term ‘*academic anxiety*’ as “*a unifying formulation for the collection of anxieties learners experience while in schools*”. While it gives the impression that anxiety shares the same nature and consequences, different types of anxieties can be triggered under different circumstances. Clement (1980:25) views foreign language anxiety as “*a complex construct that deals with learners' psychology in terms of their feelings self-esteem and self confidence*”. On their part, MacIntyre and Gardner (1994:37) describe anxiety as “*the feeling of tension and apprehension specifically associated with second language contexts, including speaking, listening, and learning*”.

Testing as a word calls into mind the feeling of constant worry, discomfort and uneasiness, in this vein Sarson & Sarson, 1990, cited in Burns (2011:213) states that, “*One of the most common anxiety types considered to be present among students, as one of the most pervasive reactions that individuals have to stress, is test anxiety*” In a

Chapter Four: Considerations and Implications for Innovation

more down to earth terms, it has been defined by Zeidnen (1998:17) as a “*set of phenomenological, psychological, and behavioral responses that accompany concern about possible negative consequences or failure of an exam or similar evaluation situations*”.

Therefore, almost everyone may experience a feeling of uneasiness and frustration when a test approaches. In fact, it is rare to find a student who does not approach the test with a degree of test anxiety. This kind of nervousness fear and discomfort happens *before, during* and sometimes *after* an exam. While it is perfectly natural to feel so, too much anxiety levels may become discouraging.

Test anxiety appears to commonly be frequent among students; it gets in the way of their studies, their psychology and their well-being as well. Too much anxiety may block students’ performance, having difficulty retaining subjects studied for the test; furthermore, it may cause a host of problems ranging from psychological to physiological; cognitive to affective as displayed in the following table:

Physical	Emotional	Behavioral	Cognitive
Headaches, Nausea	Depression	Fidgeting	‘going blank’
Body Temperature	Disappointment	Avoidance	Racing thoughts
Sweating	Uncontrollable crying	Pacing	Difficulty concentrating
Elevated heartbeat	Helplessness	Escaping	Dread

Table4.4. Test Anxiety Symptoms (Adopted from Horwitz,2002)

Anxiety seems to inhibit students’ ability to recall or retain information when asked for. It creates a mental block or a noise in the brain that impairs on his side, reasoning a reflection.

Test anxiety is one of the most common types of anxiety common among language learners. Herein, Yerkes and Dodson (1908) study the relationship between anxiety and performance. They believe that with a certain level of anxiety students may perform better; they doubt if learners will ever have enough motivation to take the test when they lack fear of failure or encouragements to perform on a test (McDonald, 2001). Therefore, they will explore their potentials, and skills. Oddly enough, if

learners have a high level of anxiety before or during a test this may inhibit their performance and may not show their abilities (McDonald, 2001).

Besides, in the search for the possible relationships between test anxiety and academic performance, Sarason (1952) developed a Test Anxiety Questionnaire for adults and a Test Anxiety Scale for children; these tests questionnaires help in assessing individual differences in test anxiety. In view of this, Zeidner (1988: 5-6) states that:

much of test anxiety research over the past half century has been conducted to help shed light on the aversive effects of test anxiety on examinee performance, and these concerns have stimulated the development of a variety of therapeutic techniques and intervention programs”

Many researches (Bailey, 1966; Horwitz, 2005; Scarason, 1996) have contributed to find out the possible causes and factors influencing test anxiety. Horwitz et al. (1986) classify foreign language anxiety into three components as shown here:



Diagram 4.11. Foreign Language Anxiety Adopted from Horwitz(1986)

4.6.2. Factors Affecting Test Anxiety

A number of factors are being herald as the major factors influencing learners' psychology in general, and testing in particular for instance, McDonald (2001) proposes the following factors:

- **Gender** is, in fact, a factor that contributes to the development of test anxiety McDonald(2001) states “ *When comparing the test anxiety levels of males and females, females consistently scored higher than males*”. Thus, due to females' willingness to develop a test anxiety at higher rates than males, “*Women are more uncomfortable and self-conscious in testing situations than men*” (Lewis & College, 1987).
- **Biology**, here, some situations seems to be due to the DNA formulation of the learners. Some learners anxiety is hereditary. Thus, knowledge of learners' medical history could be beneficial in predicting possible states of students anxiety levels.
- **Age**, according to McDonald (2001), age is an important variable in test anxiety. According to Sarrason (1966) test anxiety typically increases with age Lee (2007:11) states that “*elementary students are more likely to show physical signs while older students have behavioral symptoms of test anxiety*”.
- **Race and Culture**, studies have shown that different races affects greatly test anxiety. Spinks and Moerdyk (1980:44) discussed how “*cultural differences in anxiety scale scores depend on the fact that a given situation known to be anxiety provoking in one culture may not give rise to anxiety, or at least two very different forms of anxiety, in another culture*”. Learners coming from different backgrounds may experience higher levels of anxiety because of different stereotypes.

- **Causes of Test Anxiety**

Testing as a word call to mind varying degrees of anxiety depending on the importance of the test, supposed difficulty of the test, and the readiness for the test.

According to Young (1991:201), *“in language testing, the greater degree of student evaluation and the more unfamiliar and ambiguous the test tasks and formats, the more the learner anxiety is produced”*.

“A perceived difficult subject would elicit higher anxiety levels, and test anxiety as a psychological condition can adversely affect people in every field of life” (Cohen, 2004:116). Therefore, there are different reasons pushing students to develop test anxiety. On the one side of the ledger, students may experience anxiety when parents set unattainable goals for them so test anxiety is developed from the fear of parental rejection Sarason (1960).

Given the high pressure teachers and educationalists at large place on student to achieve a given goal, seem to be anxiety providing. On the other hand, Na (2012) believes that teachers’ sincere support and care may reduce test anxiety, so teacher-students relationship is important factor influencing learners’ psychological state. Similarly, when students begin to compare their level with other students may create certain competitiveness, this may be anxiety provoking. Anxiety can *“make learners get discouraged, lose faith in their abilities, escape from participating in classroom activities, and even give up the effort to learn a language well”* Na (2007:30).

- **Coping with Test Anxiety**

Coping with test anxiety is a challenging task that teachers, parents and stakeholders at larges are to assume psychologically and pedagogically. In order to do well in a test, students need to know well what to study and devise a plan on how to study it. Teachers must provide learners with successful testing strategies and scoring techniques; such techniques may help learners relax stay focused and become motivated to take the test. It is provided by Bass et al., (2002: 28) *“Students are more likely to be comfortable and relaxed during a test when they have greater confidence in their test taking skills”*.

Self confidence is another factor that may lessen test anxiety rates. Teachers may help learners gain confidence by teaching them memorizing techniques and skills. Time management and smart learning is also the key towards success in reducing foreign language learning anxiety.

The quest of studying the possible relationships between test anxiety and learners academic achievements does not only *"contributes to our understanding language achievement, but it is also fundamental to our understanding of how learners approach language learning, their expectations of success or failure, and ultimately why they continue or discontinue study"* Horwitz, (2001:122). Test anxiety has diverse impacts on how students input process and retain what they have learned.

There exist a number of factors that can produce anxiety inside the classroom especially in foreign language classes, among these factors the learning styles of learners and teachers strategies. In this vein, Horwitz (2001) states some tips to cope with test anxiety to insure better academic achievements these tips are summarized as follows:

- ✓ Preparation, working hard before the exam, trying to take notes and summaries of the lectures,
- ✓ Relaxation strategies, applying some breathing techniques trying to stay calm and focused,
- ✓ Positive Thinking and high self confidence,
- ✓ Peer seeking asking other apprentices if they understand the lecture,
- ✓ Resignation staying calm and giving up negative thoughts.

Research carries out on the relationship between test anxiety and learners' achievements reveal that anxiety can impede foreign language production and achievements. Aida (1994). In this sense, Oxford(1991) reported that serious language anxiety may cause problems with self esteem self confidence and risk taking abilities which will ultimately hamper proficiency in second language learning and automatically affects academic achievements.

4.7. NEW DIRECTIONS TO WASHBACK PHENOMENON

Undoubtedly, washback is a highly complex colossal phenomenon (Watanabe 2004). The fundamental focus of washback within this research work can be twofold. The first one is to describe this intricate phenomenon; what and how a speaking test affects teaching and learning and what are the factors contributing to the washback effects. Besides, the ultimate aim is to try to promote positive washback and avoid negative washback. In a more comprehensive manner, tests should be revised to generate positive washback. To reach such a challenge within a speaking test, it should be noted that test score represents a complex of multiple influences. Teachers should consider the fact that scores solely cannot be an indicator of learners' competence. What teachers are measuring can be affected by other characteristics interacting with each other, such as contents of the test, the test taker personality, learners' styles and strategies when taking a test, this will lead teachers to really test their learners in action as stated by Morrow (1986:6) who claims that *"in essence, an examination of washback validity would take testing researchers into the classroom in order to observe the effects of their tests in action"*.

Researchers often raise their concerns about whether washback is a good or a bad phenomenon (Cheng & Curtis, 2004; Wall, 1997). Views are different, some consider it a positive phenomenon, and provide good results, Wall, (1997:292) states that tests *"should be simple, its syllabus should be teachable and its effects should be beneficial"*. On the other hand, others (Madaus, 1988) consider it a negative phenomenon, Wall (1997:292) summarises Madaus' vision as follows:

The power of tests is a perceptual phenomenon, the higher the stakes attached to a test the more it will distort the teaching process, past exam papers eventually become the teaching curriculum, teachers adjust their teaching to fit the form of exam questions, test results become the major goal of schooling, and the agencies which set or control examinations eventually assume control over the curriculum.

It is noticed that almost all washback studies shed all their light on its direction (positive or negative), or if positive washback overpass negative washback. Yet, away from this vision, other measurements should be added to have a clearer image such as if the test has achieved its designed objectives and consider the time needed for washback effects. Notwithstanding the various proposed models of washback, it seems to be impossible to have a model covering all its aspects.

4.8. CONCLUSION

This part of the research work aimed at suggesting solutions to promote positive washback in foreign language assessment. In the course of these prepositions, students' awareness may be raised about the impact of tests have on their performance and their achievements. In addition to this, it proposed a number of recommended solutions and new testing paradigms to be used by teachers in the process of language testing.



General Conclusion

New challenges engendered by changes in the world economic, social, and educational contexts of the 21st century assure that language teaching/assessing will be vibrant and exciting. Continuing issues include the test validity of scoring techniques, and more importantly assessing professionally is in vogue within our research work. Our main challenges within this research include designing classroom language assessment, helping teachers be good test designers/evaluators, to hopefully develop teachers' assessment professionalism.

It is a truth generally acknowledged that the results of some standardized tests seem to be so important for students and for educational programs as well. There are clear indicators that the student is having problems if he/she fails to get good results and encounter problems while being tested, and eventually fails with low scores on tests. This appears to be preventing him from being an active performer, ready to overcome anything that gets into his/her way towards success.

Hence, in order to know the reasons of low achievement rates within learners' oral proficiency, the researcher conducted an action research to demonstrate the ambiguity of washback phenomenon on the teaching strategies adopted by the teachers to develop learners' oral proficiency. It also sought to measure the correlation between teaching-to-the-test, washback and learners' oral proficiency.

To achieve a well presented research work, four chapters were presented. The first chapter was about the literature review, including previous researches in the area, terminology description and the theoretical basis of this research work. The second chapter aimed at giving a brief description of the data collection procedure, description of the research instruments employed and the sampling management. As for the third

General Conclusion

chapter, it was concerned with the data analysis and interpretation of the results achieved from the data collection phase. Coming to the fourth chapter, a set of suggestions and recommendations were administered to both teachers and learners for better teaching/testing experience.


Results obtained were interesting, they show that there is a significant relationship between learners' oral proficiency, washback and learners' test scores. In fact, results demonstrate that there is a statistical association between the teaching method followed by the teacher and learners' oral proficiency. The more the test has a negative washback on teaching the more the oral proficiency decreases. The more the test has a positive washback on teaching, the more proficient speakers they become. Hence, when teachers follow an ordinary teaching not exam- related instruction, student' oral proficiency increases, and their speaking achievement increases and vice versa.

Besides, our findings also denote that there is a strong positive statistical relationship between test washback and learners' test scores. Thus, as the test has a high impact on teaching, the scores are influenced too. Therefore, there is a *direct proportion* between the two variables. Oddly enough, the researcher discovers from the results achieved that there is no statistical association between the oral proficiency and the test scores.

However, a number of limitations were encountered; first the study was limited to one group of second year LMD students at Tlemcen University, and the sample size was too small for generalizability. Moreover, the researcher had little knowledge about statistical package SPSS, for this reason, the researcher enrolled into SPSS training in Sidi Belabes and completed a training of 18hours (see Appendix P). Furthermore, the researcher could not get access to speechace premium due to its very expensive charges (see Appendix 'Q') though it proves very interesting support to assess learners' oral proficiency in less time, less energy, and with certain accuracy.

General Conclusion

It is expected that this study would contribute to the world of testing in general, and it would help teachers become professional language designers/testers. However, further studies are required to explore deeply the washback phenomenon from a more distinctive angle. In this sense, one may wonder whether teachers can opt for Web-Based Assessment in High-Stakes testing and avoid negative washback and how do the teacher's backgrounds and perceptions influence the way that the courses are delivered/tested? These questions may open the doors for further reading and future research.



Bibliography

Bibliography

- Agrawal, M. (2004) 'Curricular reform in schools: the importance of evaluation', *Journal of Curriculum Studies*.
- Aida, Y. (1994). Examination of Horwitz, and Cope's construct of foreign language anxiety. The case of students of Japanese. *Students. Modern Language Journal*
- Akanbi, S. T. (2010). Test anxiety as a correlate of academic achievement among senior secondary school students in Ogbomoso area of Oyo State. *African Journal of Educational Research*
- Alderson, J. C., Clapham, C. & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Alderson, P. & Morrow, V. (2004). *Ethics, social research and consulting with children and young people*. London: Barnardos
- Amengual-Pizarro, M. (2009) 'Does the English Test in the Spanish University Entrance Examination Influence the Teaching of English?', *English studies*
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders (3rd ed.)*. Washington, DC: Author.
- American Psychological Association, (2013) *Publication manual of the American Psychological Association sixth edition*. Washington, DC: APA
- Anderson, S. (1993). Trouble with testing. *The American School Board Journal*, 24-26.

Bibliography

- Andrews, S. (1994). The Washback Effect of Examination: Its Impact Upon Curriculum Innovation in English Language Teaching. *Curriculum Forum* 4(1), 44-58.
- Andrews, S. (2004). Washback and Curriculum Innovation. In CHENG, L. et al. *Washback in Language Testing: Research Contexts and Methods* LEA Publishers London
- Ayers, W. (1993). *To teach: The journey of a teacher*. New York: Teachers College Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press
- Bailey, K. (1983). "Competitiveness and Anxiety in Adult Second Language Learning: Looking at and through the Diary Studies" in Brown, , D. (1994), *Principles of Language Teaching and Learning* . 3 rd ed. New Jersey: Prentice Hall, Inc.
- Bailey, K.M. (1996). Working for Washback: A Review of the Washback concept in Language Testing. *Language Testing*
- Baker, C. (1992) *Attitudes and language*, Clevedon: Multilingual Matters Ltd.
- Barnes, M. and Clarke, D. and Stephens, M. (2000) 'Assessment: the engine of systemic curricula reform?' *Journal of Curriculum Studies*.
- 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, Tlemcen.
- Benmoussat, S. (2003). *Mediating Language and Culture: An investigative Analysis of the Cultural Dimension in the Algerian Newly-Designed ELT Textbook*. Unpublished Doctorat d'Etat Thesis, Abou Bakr Belkaid University, Tlemcen.
- Benmoussat, S. (2018). *The Teach-to-the-Test Approach: A Curse a Blessing or a Blessing in Disguise for Algerian EFL Students* in *English Language and Literature Studies*; Vol. 8, No. 3; 2018 ISSN 1925-4768 E-ISSN 1925-4776

Bibliography

- Published by Canadian Center of Science and Education. Retrieved from <http://www.ccsenet.org/journal/index.php/ells/article/download/73765/41730>
- Bhat (2015) *Types and Methods of Interviews in Research*. retrieved from <https://www.questionpro.com/blog/types-of-interviews/>
 - Bhat, A., 2015, *Types and Methods of Interviews in Research*, retrieved from: <https://www.questionpro.com/blog/types-of-interviews/>
 - Black, S. (2005). Test anxiety. *American School Board Journal*, 192(6), 42-44.
 - Bouhadiba, Z. (2013). THE LMD SYSTEM IN ALGERIA: THE CASE OF ENGLISH. 7th International Technology, Education and Development Conference. Valencia, Spain. Retrieved from <https://library.iated.org/view/BOUHADIBA2013LMD#:~:text=The%20LMD%20reform%20was%20launched,the%20academic%20year%202003%2D2004.&text=This%20paved%20the%20way%20to,of%20the%20new%20educatio%20reform.>
 - Brindley , G. (2008). Educational Reform and Language Testing. In HORNBERGER, N.H (Ed.) (2008) *Encyclopedia of Language and Education*, 2nd Edition Springer.
 - British Council (2018) Technical Report Aptis General Technical Manual Version 1.0:
 - Brown, D. H. 2001. *Teaching by Principles: An Interactive Approach to Language Pedagogy*. (2nd Edition). Addison Wesley Longman, New York.
 - Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. New York: Longman.
 - Brown, H.D. and Abeywickrama, P. (2010) *Language assessment : principles and classroom practices*, 2nd ed. Upper Saddle River, N.J. ; Harlow: Pearson Education
 - Brown, J. D.1998. *Understanding Research in Second Language Learning*. Cambridge, UK, Cambridge University Press
 - Burns, A. (1999). *Collaborative action research for English language teachers* Cambridge: Cambridge University Press.

Bibliography

- Burns, A. (2010). *Doing action research for English language teachers: A Guide for Practitioners*, Cambridge: Cambridge University Press.
- Byman, A., & Cramer, D. (1999). *Quantitative Data Analysis With SPSS Release for Windows : a guide for social scientists*. London: Routledge.
- Canale, M. and SWAIN, M. (1982). *Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing*. Applied Linguistics
- Canale, M. (1983). On Some Dimensions of Language Proficiency. In J.W. Oller, ed., *Issues in Language Testing Research*. Rowley, MA: Newbury House, 333-342.
- Canale, M. and SWAIN, M. (1980). *Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing*. Washington, D.C. Center of Applied Linguistics.
- Canale, M. and SWAIN, M. (1980). *Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing*. Washington, D.C. Center of Applied Linguistics.
- Canale, M. and SWAIN, M. (1982). *Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing*. Applied Linguistics
- Carrol, J. B. 1965. "Fundamental Consideration in Testing for English Language Proficiency of Foreign Students". In Allen, J. P. B. (ed.) *Teaching English As a Second Languages*. New York: McGraw Hill, Inc
- Cassady, J. C. (Ed.). (2010). *Anxiety in schools: The causes, consequences, and solutions for academic anxieties*. New York: Peter Lang.
- Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27(2), 270-295. doi:10.1006/ceps.2001.1094
- Cattell, R. B., & Scheier, I. H. (1961). *The meaning and measurement of neuroticism and anxiety*. Oxford, England: Ronald Press Co.
- Chapman, D. W. and Snyder Jr, C. W. (2000) 'Can high stakes national testing improve instruction: reexamining conventional wisdom', *International Journal of Educational Development*, 20(6), 457-474

Bibliography

- Cheek, J. R., Bradley, L. J., Reynolds, J., & Coy, D. (2002). An intervention for helping elementary students reduces test anxiety. *Professional School Counseling, 6*(2), 162-165.
- Cheng, Y.-S., Horwitz, E. K., & Schallert, D. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning, 49*, 417–446.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*
- Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*
- Cheng, L. (1999). Changing assessment: Washback on teacher perspectives and actions. *Teaching and Teacher Education*
- Cheng, L. (2002). The washback effect on classroom teaching of changes in public examinations. In Savignon, S. J. (eds.), *Interpreting communicative language teaching: Contexts and concerns in teacher education*. New Haven, CT: Yale University Press
- Cheng, L. (2005). Changing language teaching through language testing: A washback study. *Studies in Language Testing 21*. Cambridge: Cambridge University Press and Cambridge ESOL
- Cheng, L. (2008). Washback, impact and consequences. In Shohamy, E. & Hornberger, N. H. (eds.), *Encyclopedia of Language and Education* (2nd edn). *Language Testing and Assessment* (vol. 7), New York, NY: Springer,
- Cheng, L., Watanabe, Y. & Curtis, A. (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates
- Chomsky, N. (2015) The Danger of Standardized Testing <https://creativesystemsthinking.wordpress.com/2015/02/21/noam-chomsky-on-the-dangers-of-standardized-testing/>

Bibliography

- Clark, V.L. Gutmann, M.L. and Hanson, W.E. (2003) 'Advanced mixed methods research designs', Handbook of mixed methods in social and behavioral research.
- Clément, R. (1980). Ethnicity, contact, and communicative competence in a second language. New York, NY: Springer,
- Clément, R., & Gardner, R. C. 2001. Second Language Mastery. In H. Giles & W. P. Robinson (Eds.), *The New Handbook of Language and Social Psychology* (2nd ed.). London: Wiley.
- Clément, R., & Kruidenier, B. (1983). Orientations on second language acquisition: 1. The effects of ethnicity, milieu and their target language on their emergence. *Language Learning*.
- Clément, R., Dörnyei, Z., & Noels, K. 1994. 'Motivation, Self-Confidence and Group Cohesion in the Foreign Language Classroom'. *Language Learning*.
- Cohen, D 1989, 'Tests Fail to Pass Their Exams', in G Shrubbs (ed), *Mass Testing Tonic or Toxin? Counterpoints No. 1*, Australian Education Network.
- Cohen, L., Manion, L. & Morrison, K. R. B. 2007. *Research Methods in Education*, New York, Routledge.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*
- Deffenbacher, J. L. (1986). Cognitive and physiological components of test anxiety in real-life exams. *Cognitive Therapy and Research*
- 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 of Tlemcen. (Doctoral Thesis). Tlemcen
- Djekoun, Abdelhamid (2006). *La réforme LMD en Algérie*, Retrieved from <http://www.umc.edu.dz/VersionFrancais/menu%20vertical/systeme%20lmd/systemeLmdAlgerie.htm#preambule>, accessed December 4, 2009.

Bibliography

- Dörnyei, Z. & Skehan, P., 2003. Individual differences in second language learning. In: DOUGHTY, C. J., LONG, M. H., eds. *The Handbook of Second Language Acquisition*. Oxford : Blackwell, pp. 589-630
- Dörnyei, Z. (2003). Attitudes, orientations, and motivations in language learning: Advances in theory, research and applications. *Language Learning*.
- Dörnyei, Z. (2003). Questionnaires in Second Language Research: Construction, Administration, and Processing. Mahway, NJ: Lawrence Erlbaum.
- Dörnyei, Z. (2005). The psychology of the language learner: Individual differences in second language acquisition. Mahwah, USA: Lawrence Erlbaum Associates.
- Dörnyei, Z. 2001. *Teaching and Researching Motivation*, Pearson Education, New York.
- Fatemeh Alipanahi. 2016. Formative Assessment, Reading Comprehension and EFL Learners Autonomy *Indian Journal of Fundamental and Applied Life Sciences* ISSN: 2231 6345
- Ferman, I. (2004) 'The washback of an EFL national oral matriculation test to teaching and learning', in Cheng, L. Watanabe, Y., and Curtis (ed.) *Washback in References 286 Language Testing: Research Context and Methods*, Mahwah New Jersey USA: Lawrence Erlbaum Associates.
- Finch, A. E. & Shin, D-I, (2005). Integrating teaching and assessment in the EFL classroom: A practical guide for teachers in Korea . Seoul: Sahoipyungnon Publishing Co., Inc
- Flavell, J. H. (1981) Cognitive Monitoring. In W. P. Dickson (Ed.), *Children's Oral Communication* (pp. 35-60). New York: Academic Press.
- Fredericksen, J.R. and Collins, A. (1989) 'A system approach to educational testing', *Educational Researcher*
- Fulcher, G. and Davidson, F. (2012) *The Routledge handbook of language testing*, London: Routledge.

Bibliography

- Gardner, R. C (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.
- Garrett, H. E. (1964). *Statistics in psychology and education*. New York: Longmans, Green
- Gates (1995). *Language learning and perfectionism: Anxious and non-anxious language learner's reactions to their own oral performance*. *Modern Language*.
- Gorsuch, G.J. (1999) *Exploring the relationship between educational policy and instruction in Japanese high schools EFL classrooms*, unpublished PhD thesis. The Temple University: Japan.
- Gorsuch, G.J. (1999) *Exploring the relationship between educational policy and instruction in Japanese high schools EFL classrooms*, unpublished PhD thesis. The Temple University: Japan.
- Gorsuch, G.J. (2000) 'Educational policies and educational cultures: influences on teachers' approval of communicative activities.', *TESOL Quarterly*
- Gorsuch, G.J. (2000) 'Educational policies and educational cultures: influences on teachers' approval of communicative activities.', *TESOL Quarterly*
- Gregory, C. and Burg, S. (2006) *Addressing test anxiety in high-stakes environment: Strategies for classrooms and schools*, London: Sage
- Haladyna, T.M. Nolen, S.B. and Haas, N.S. (1991) 'Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution', *Educational Researcher*
- Harri-Augstein S., & Thomas, L. (1991). *Learning conversations: The self-organised learning way to personal and organisational growth*. London: Routledge.
- Heng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In Cheng, L., Watanabe, Y. & Curtis, A. (eds.), *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bibliography

- Heng, L., Andrews, S. & Yu, Y. (2011). Impact and consequences of school-based assessment in Hong Kong: Views from students and their parents. *Language Testing*
- Herman, J., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice*,
- Herman, J.L. (1992). Complex PTSD: A syndrome in survivors of prolonged and repeated trauma. *Journal of Traumatic Stress*, Herman, J.L.,
- Horwitz, E. K. , Horwitz M. B. and Cope, J. (1986). “Foreign Language Classroom Anxiety ” *The Modern Language Journal*,
- Horwitz, E.K. (1986). Preliminary evidence for the reliability and validity of a foreign language anxiety scale. *TESOL Quarterly*, 20, 559–562.
- Horwitz, E.K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*.
- Hsu, H. (2009) The impact of implementing English proficiency tests as a graduation requirement at Taiwanese universities of technology, unpublished PhD thesis. University of York.
- <https://www.questia.com/library/journal/1P3-4309525071/factors-influencing-test-anxiety-among-university>
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hughes, A. (2003) *Testing for language teachers*, 2nd ed. Cambridge: Cambridge University Press.
- Hughes, Arthur. (1995). *Testing for Language Teachers* . Seventh Printing. University of Cambridge. Bell &Bain, Ltd. Glasgow
- Isaacs, T. (2016). Assessing speaking. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 131–146). Berlin: DeGruyter Mouton. 2019, From https://www.researchgate.net/profile/Talia_Isaacs/publication/298213995_Assessing_speaking/links/5884ea2b4585150dde485d75/Assessing-speaking.pdf

Bibliography

- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test* (2nd ed.). Circle Pines, MN: American Guidance Service.
- Kaufman, D. (2004). The Negative Effects of High-Stakes Testing. Pamela Spann Education Law and Policy
- Kemmis, McTiggart. (1986) *The Action Research Planner: +Business Media* Singapore. Retrieved from <https://link.springer.com/content/pdf/10.1007%2F978-981-4560-67-2.pdf>
- Kemmis, Stephen and McTaggart, Robin (1986) *The Action Research Planner*, 3rd edition. Geelong, Victoria: Deakin University Press.
- Kitao, S. K., & Kitao, K. (1996). Testing communicative competence (Report No. TM025214). (ERIC Document Reproduction Service No. ED398260)
- Kitao, S. K., & Kitao, K. (1996). Testing speaking (Report No. TM025215). (ERIC Document Reproduction Service No. ED398261)
- Lado, Robert. (1975). *Language Testing*. Wing Tasi Cheung Printing Company.
- Lam, H. (1995) 'Methodology washback-an insider's view'. in D. Nunan R. Berry and V. Berry (eds.) *Bringing about change in language education*, Hong Long: University of Hong Kong.
- LeCompte, M. D Schensul (1999) *Analyzing and Interpreting Ethnographic Data*, Ethnographic toolkit 5. California: Alta Mira Press
- Liepmann, D., Marggraf, C., Felfe, J., & Hosemann, A. (1992). Anxiety, action orientation, subjective state and situational aspects: A study of tank-lorry drivers. In K. Hagtvet & T. B. Johnson (Eds.). *Advances in test anxiety research* (Vol. 7, pp. 130-141). Amsterdam/Lisse, The Netherlands: Swets & Zeitlinger.
- Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2018). Paradigmatic controversies, contradictions, and emerging confluences revisited. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* Thousand Oaks, CA: Sage.

Bibliography

- MacIntyre, P.D., & Gardner, R.C. (1989). Anxiety and second language learning: Toward a theoretical clarification. *Language Learning*, 32, 251-275.
- Madaus, G. (1988) 'The influence of testing on the curriculum'. in L. Tanner (ed.) *Critical issues in curriculum*, University of Chicago Press
- Madsen, Horald S, 1987. *Techniques in Testing*. Oxford: OUP
- Madsen, Horald S., 1981. "Selecting Appropriate Elicitation Techniques for Oral Proficiency Tests" in John A.S. Read. *Directions in Language Testing*. Singapore: RELC, pp. 87-99.
- McDonald, A. D. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology*, 21, 89 – 101.
- McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 21(1), 89-101. doi: 10.1080/01443410020019867
- McNamara, Carter, PhD. *General Guidelines for Conducting Interviews*, Minnesota, 1999
- McNamara, Tim. (2001). *Language Testing*. Oxford University Press
- Messick, S. (1996) 'Validity and washback in language testing.', *Language Testing*
- Meziane, M. K. (2010). *Language and Gender: A Study of Interruptions as a Cause of Conflicts in Mixed-Sex Conversations (Unpublished Magister thesis in Sociolinguistics)*. Abou Bekr Belkaid University of Tlemcen, Algeria.
- Miliani, M., (2003). *Foreign Language Teaching Approaches, Methods and Techniques*. Oran: Dar El Gharb.
- Miliani, M., (2010). *Between Enduring Hardships and Fleeting Ideals. Mediterranean Journal of Educational Studies, Vol. 15(2), pp. 65-76, 2010*
- Moerdyk (1980). The relationship of high school graduation exams to graduation rates and SAT scores. *Education Policy Analysis Archives*,

Bibliography

- Mohammed, M., & Brahim, M. (2010). The LMD higher education system in the Maghreb countries: The example of Algeria. In B. Lamine (Ed.), *Towards an Arab higher education space: International challenges and societal responsibilities* Beirut: UNESCO Regional Bureau for Education in the Arab States.
- Morrow, Keith, 1982. "Testing Spoken Language" in JB Heaton (ed.). *Language Testing*. Great Britain: Modern British Publication
- Morrow, Keith, 1982. "Testing Spoken Language" in JB Heaton (ed.). *Language Testing*. Great Britain: Modern British Publication Ltd, Pp. 56-58.
- Na Z (2007). A study of high school students' English learning anxiety. *Asian EFL J.*
- Nichols, S. L. & Berliner, D. C. (2005). The inevitable corruption of indicators and educators through high-stakes testing. *Education Policy Studies Laboratory*. Retrieved July 2 2018 from <http://edpolicylab.org/>.
- Nie, N. H. (1975). *SPSS: Statistical Package for Social Sciences*. New York: McGraw-Hill
- Nieuwenhuis S, Jepma M, Cohen JD (2010) Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn Affect Behav Neurosci.*
- Odo, D.M. (2012) 'The impact of high school exit exams on ESL learners in British Columbia', *English Language Teaching*, 5(9), 1-8.
- Odo, S. 2012. *What Do Test Scores in Texas Tell Us?* Santa Monica, CA: RAND
- Orafi, S.M.S. (2008) Investigating teachers' practices and beliefs in relation to curriculum innovation in English language teaching in Libya, Unpublished PhD thesis. University of Leeds.
- Oxford, R. 1990. *Language Learning Strategies :What every Teacher Should Know*. Oxford ; Heinle and H , publication.

Bibliography

- Oxford, R. L. (2005). Anxiety and the language learner: New insights. In J. Arnold (Ed.), *Affect in language learning* (p. 58). Cambridge, MA: Cambridge University Press.
- Oxford, R. 1994 "Language Learning Strategies: An Update" in *ERIC Digests*: ERIC Clearinghouse on Languages and Linguistics Washington
- Peacock, A.T. & Wiseman, J. (1961). *The Growth of Public Expenditure in the United Kingdom*, London: Oxford University Press.
- Pearson (2020, February) Online Course Toolkit Digital Solutions. Pearson USA Retrieved from <https://www.pearson.com/us/campaigns/online-course-toolkit/digital-solutions.html>
- Pearson, I. (1988). Tests as levers for change. In D. Chamberlain & R. J. Baumgardner (Eds.), *ESPM in the classroom: Practice and evaluation* (pp. 98-107). London: Modern English
- Peirce, ML (1966). The subjective experience of FL anxiety: Interviews with highly anxious students.
- Petrie, H. (1997). User-centred design and evaluation of adaptive and assistive technology for disabled and elderly users. *Informationstechnik und Technische Informatik*,
- Popham, W.J. (1987) 'The merits of measurement-driven instruction', *Phi Delta Kappa*..
- Popham, W.J. (1991) 'Appropriateness of teachers' test-preparation practices', *Educational Measurement: Issues and Practice*,
- Qi, L. (2004) 'Has a high-stakes test produced the intended changes'. in L. Cheng, Y. Watanabe and A. Curtis (ed.) *Washback in language testing: Research context and methods*.
- Qi, L. (2004) 'Has a high-stakes test produced the intended changes'. in L. Cheng, Y. Watanabe and A. Curtis (ed.) *Washback in language testing: Research context and methods*. 171-190.

Bibliography

- Qi, L. (2007) 'Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes test in China', *Assessment in Education: Principles, Policy & Practice*,
- Qian, David D. (1998) *Depth of vocabulary knowledge: assessing its role in adults' reading comprehension in English as a second language*. Ontario Institute for Studies in Education, University of Toronto.
- Robinson, & P. M. Smith (Eds.), *Language: Social psychological perspectives*. Oxford, United Kingdom: Pergamon Press. pp. 147-154
- Sarason, I. G. (1984). *Stress, anxiety and cognitive interferences: Reactions to tests*. *Journal of Abnormal and Social Psychology*.
- Sarason, I. G. (1980) *Test Anxiety: Theory, research, and application*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sarason, I. G., & Stoops, R. (1978). *Test anxiety and the passage of time*. *Journal of Consulting and Clinical Psychology*,
- Shepard, L.A. (1993) 'The place of testing reform in educational reform: A reply to Cizek.', *Educational Researcher*
- Shohamy, E. (1992). *Beyond Proficiency Testing: A Diagnostic feedback Testing Model for Assessing Foreign Language Learning*. *Modern Language Journal* 76(4) 513-521.
- Shohamy, E. (1996). *The Power of Tests: The Impact of Language Tests on Teaching and Learning*. Washington, D.C.: National Foreign Language Center.
- Shohamy, E. (1992) 'Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning', *The Modern Language Journal*, 76(4), 513-521.
- Shohamy, E. (1993) 'The power of tests: The impact of language tests on teaching and learning.'. NFLC Occasional Paper. Washington: National Foreign Language Center.
- Shohamy, E. (2001) *The power of tests : a critical perspective on the uses of language tests*, Harlow: Longman.

Bibliography

- Shohamy, E. (2007) 'Tests as Power Tools'. in J. Fox M. Wesche D. Bayliss L. Cheng C. Turner and C. Doe (eds.) Language Testing Reconsidered, Ottawa: University of Ottawa Press
- Smith, J.K. (1983) 'Quantitative versus qualitative research: An attempt to clarify the issue', Educational Researcher
- Spielberger. Theory and research on anxiety. In : Anxiety and theory. SPIELBERGER CD (ed). Academic Presse, New York, 1966.
- Spielberger, C. D. (1979). Understanding stress and anxiety. New York: Harper & Row
- Spolsky, B. (1995) 'The examination of classroom backwash cycle: Some historical cases', in D. Nunan V. Berry and R. Berry (eds.) Bringing about change in language education, Hong Kong: University of Hong Kong, Department of Curriculum Studies
- Spratt, M. (2005) 'Washback and the classroom: the implications for teaching and learning of studies of washback from exams', Language Teaching Research, 9(1), 5- 29.
- Stecher, B. (2002) 'Consequences of large-scales, high-stakes testing on schools and classroom practices.', in L. Hamilton, B. Stecher and S. Klein (ed.) Making Sense of Test-Based Accountability in Education, CA: Rand, USA.: Santa Monica,
- Stecher, B. Chun, T. and Barron, S. (2004) 'The effects of assessment-driven reform on the teaching of writing in Washington State. ', in L. Cheng, Y. Watanabe and A. Curtis (ed.) Washback in Language Testing: Research Context and Methods, Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc
- Sujana, M. (2016): Assessing Oral Proficiency: Problems And Suggestions For Elicitation Techniques, retrieved from: <https://imadesujana.wordpress.com/2016/09/07/assessing-oral-proficiency-problems-and-suggestions-for-elicitation-techniques/>

Bibliography

- The British Council. (2019, March 12). General technical Manual. The British Council. Retrieved on 21 October 2019, from https://www.britishcouncil.org/sites/default/files/aptis_general_technical_manual_v-1.0.pdf
- Tsagari, D. (2009a) 'Revisiting the concept of test washback: Investigating FCE in Greek Language schools'. 35. Cambridge: Cambridge ESOL.
- Tsagris, M. T, (2009). A data-based power transformation for compositional data. In Proceedings of the 4th Compositional Data Analysis Workshop, Girona, Spain.
- Valerie Strauss, *The brainy questions on Finland's only high-stakes standardized test*, WASHINGTONPOST.COM, (March 24, 2014)
- Valette, R.M. (1967). Modern Language Testing. New York: Harcourt Brace.
- Valette, R.M. (1977). Modern Language Testing. 2nd Ed. New York: Harcourt Brace.
- Vollmer, T. R., & Bourret, J. (2000). An application of the matching law to evaluate the allocation of two- and three-point shots by college basketball players. *Journal of Applied Behavior Analysis*.
- Wall, D and ALDERSON, J.C. (1993). Examining Washback: The Sri Lanka Impact Study. In A. CUMMING and R. BERWICK (Eds) *Validation in Language Testing* (194-221). Philadelphia: Multilingual Matters. WALL, D. (1996). Introducing New Tests into Traditional Systems: Insights from General Education and from Innovation Theory. *Language Testing*
- Wall, D. (1996) 'Introducing new tests into traditional systems: Insights from general education and from innovation theory.', *Language Testing*.
- Wall, D. (1997) 'Impact and washback in language testing', *Encyclopaedia of language and education*, 7 291-302.
- Watanabe, Y. (1996) 'Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research.', *Language Testing*, 13(3), 318-333. Watanabe, Y. (2000) 'Washback effects of the English

Bibliography

- section of the Japanese university entrance examinations on instruction in pre-college level EFL', *Language Testing Update*, 27 42-47.
- Watanabe, Y. (2004) 'Methodology in Washback Studies'. in A. Curtis (ed.) *Washback in language testing: Research contexts and methods*. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers,
 - Watanabe, Y. (2004). *Teacher factors mediating washback*. New York, NY: Springer.
 - William A. Firestone, , Roberta Y. Schorr, Lora F. Monfils (2004) *The Ambiguity of Teaching to the Test: Standards, Assessment, and Educational Reform*. Lawrence Erlbaum Associates, Mahwah, NJ
 - Xi, X. (2010b) 'Automated scoring and feedback systems: Where are we and where are we heading?', *Language Testing*
 - Young DJ (1991). Creating a low-anxiety classroom environment: What does the language anxiety research suggest? *Modern Lang. J.* 75:425-439.
 - Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum Press.
 - Zheng Y (2008). Anxiety and second/foreign language learning. *Canadian J. New Scholars Educ.*





Appendix –A-
Needs Assessment Questionnaire

Needs Assessment Questionnaire

1. According to you, what is the importance of the oral production course?
.....
.....
.....
2. Describe your previous experience in the oral course
.....
.....
3. What speaking difficulties do you face in performing an oral task?.....
.....
.....
4. According to you, which part of speaking would you like to improve?
 - Grammar
 - Vocabulary
 - Fluency
 - pronunciation
5. What activities do you think will help you improve your speaking proficiency level?
.....
.....
6. Think back to a useful and enjoyable training course you attended. What made that course so good for you?
.....
.....
.....
7. What expectations do you have of your teacher and the oral production module?
.....

Appendices

-
-
8. Is there anything else we should know about? For example, special needs, learning difficulties, or other factors which might affect your learning
-

✓ Read each statement carefully. If the statement reflects your experience in taking a test, please place a ✓ in the appropriate box strongly agree, agree; strongly disagree, disagree.

Item	Statement	Strongly Agree	Agree	Neutral	Strongly Disagree	Disagree
1.	I can express ideas and opinions with very high level of fluency.					
2.	I can Show very high level of comprehension, confidence and accuracy in speaking.					
3.	I display high levels of critical thinking					
4.	I can interact effectively in speaking English					
5.	I am able to present information in sequence and interact accurately					
6.	I am able to understand idioms and various meanings of words					
7.	I am able to convey the message according to the intention.					

We will treat the information you have given with the strictest confidence, using it

- a) To design suitable courses for you and,
b) To personalize materials for your course to meet your needs.*

Thank you very much for your collaboration



Appendix B
Washback Questionnaire

Appendix B

WASHBACK QUESTIONNAIRE I

Dear teachers, you are kindly requested to put a tick (√) on the statement that best describes your opinion.

General information

- Your Teaching Experience:
- Modules in charge of:.....
- Post graduate Degree:.....

N	Statement	Strongly Agree	Agree	Strongly Disagree	Disagree	Not Sure
1	Examinations influence my teaching of speaking.					
2	The courses I design for my students help them get prepared for examinations.					
3	I use hidden curriculum in teaching speaking.					
4	I don't teach my students according to the prescribed syllabus.					
5	Examinations oblige me to teach selected topics.					
6	I feel anxious to bring good results in the oral tests.					
7	The current course help learners improve their oral proficiency					
8	Teaching test-taking techniques is most important in my class					
9	I design my oral tests according to what I have taught in the classroom					
10	Students' performance on the test reflects their abilities					
11	Examinations affect my teaching process every year.					

Thank you very much for your collaboration



Appendix C
Learners' Oral proficiency Test

Appendices

English Department

University Year 2018/2019

Second year Oral Proficiency Test
Module : Oral Production

LMD2 G1
Semestre 1

Lab.1.

ORAL PROFICIENCY TEST

QUESTIONS

Part One.

Q1. Identify the meanings:

Idiomatic Expression	Idiomatic Expression
What's eating you? Bite the bullet, in this neck of the woods, the bottom line, flunk a test, have got it made, have no clue, beyond the shadow of a doubt	Out of the blue Split milk Climb the walls Be knocked out Keep a level head Sit tight The last straw

Q2. Using the idiomatic expression learned answer in less than a minute

1. Tell us about a film you really like
2. Are you doing anything special this weekend
3. What acts of kindness have you performed?

Part Two: (recorded)

Q1. Talk about the photograph (one minute)

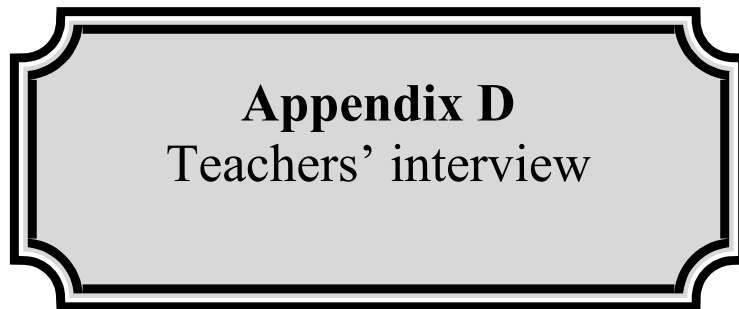
Two people helping other people in different situations.

- Compare the photograph and say how important it is to help people.

Q2. Topic: (Three minutes nonstop talk)

Phobias are strong fears; "claustrophobia" is the fear of being stuck in a small space, such as an elevator. Do you or anyone you know have any phobia?

Good luck!



Appendix D
Teachers' interview

Teachers' interview

- Gender:.....
- Your Teaching Experience:
- Modules in charge of:.....
- Post graduate Option:.....
- Post graduate Degree:.....

Rubric One : Oral Proficiency

1. What do you think about second year learners' oral proficiency level?
 - Excellent
 - Good
 - Fair
 - Poor
 - Very poor
2. Do you follow any specific program in teaching oral production?
 - Yes
 - No
3. Do you think that your designed course is helping improving learners' oral proficiency
How?
4. According to you, what are the learners' oral proficiency lacks?
5. Do you include oral proficiency training in your oral course?
 - Yes
 - No
6. When teaching, do you focus more on :
 - Learners' oral proficiency

Appendices

- The appropriate teaching methodology
- Course Objectives
- Test objectives

Rubric Two: Course Design

7. What is the first step you take in designing your oral course?
8. What criteria do you take into consideration when designing your course?
9. According to you, what are your learners' needs in speaking?
10. Which of the phases do you find challenging?
 - Course Design
 - Test Design
 - Scoring Phase

Why?

11. In choosing materials for class, do you choose test-related materials? Or is test-related material not a major consideration for designing your course?.
12. Do you follow any specific program in teaching oral production?
 - Yes
 - No
13. Do you think that the present course is helping improving learners' oral proficiency? How
14. What are the most significant changes you have made in your teaching within the context of CLT?
 - Teaching according to the test content
 - Adopting new teaching methods
 - Using a more communicative approach in teaching
 - Emphasizing on listening and speaking skills
 - The amount and frequency of practicing activities similar to the final exam increased in my classroom as the exam approached.

Rubric Three: Test Design

15. In assessing, on what criteria do you focus the most?
- Accent
 - Comprehension
 - Vocabulary
 - fluency
 - Grammar
 - Pronunciation
16. What is your assessment strategy used in designing a speaking test?
17. What are the ways that you follow in addressing the test?
18. What percentage of class time is devoted to test-related instruction or practice
19. Do you devote a major portion of your class time to practice test-preparation materials
20. What kind of activities do you chose in designing your oral test
21. Do you test items you have already dealt with in the classroom or you prefer challenging questions?
- Yes
 - No
22. What measures do you want your administration to provide you to reach the goals set at the onset of teaching?
23. What positive or negative effects of oral tests have been brought about on your teaching

Thank you very much for your collaboration



Appendix E
Observation for the Teacher

Appendices

OBSERVATION GRID PHASE

Dear teachers, you are kindly requested to put a tick (√) on the statement that best describes your opinion.

N	Statement	Teachers' Comment
1	The teacher designed her course based on students needs	
2	The course was designed to prepare students to the test.	
3	The course was explored enough for students to let them pass the test without preparing the course	
4	The course designed for students helps them get prepared for examinations	
5	The teacher searches only for correct answers, no further discussion.	
6	The teacher focuses on exam-related content.	
7	The teacher does not use challenging topics that have no relation with the test.	
9	Half of the class time is devoted to exam-related instruction.	
10	The teacher encourages dialogue memorizations.	
11	The teacher prepares her students to get good marks in examinations only.	

Thank you very much for your collaboration



Appendix F
Oral Proficiency Rating
Scale

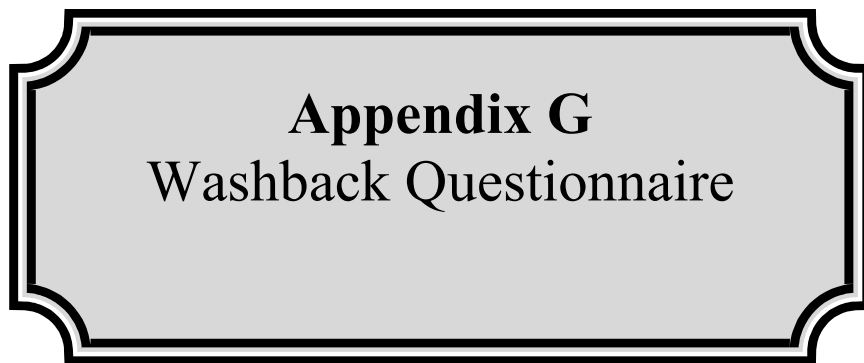
Appendices

PROFICIENCY QUESTIONNAIRE.

Dear teacher, you are kindly requested to put a tick (√) on the statement that best describe the students' proficiency.

	Statement	All the time	Most of the time	Occasio	Rarely	Not At all
1	Comprehension: Demonstrates poor and little Mastery of the topic					
2	Fluency : Control of pace is poor					
3	Vocabulary: Ineffective oral communication due grammar mistakes, incoherence					
4	Pronunciation: Uses unclear pronunciation					
5	Accuracy: Makes inaccurate definitions of concepts					
6	Grammar: Speech full of grammar mistakes					
7	show frequent long pauses, repetition of teacher's words,					
8	Oral production consists of isolated words and perhaps a few high-frequency phrases					
9	able to satisfy the necessities of a basic communication retrieving learned utterances					
10	Able to handle successfully most uncomplicated communicative tasks and social situations					

Thank You Very Much Indeed



Appendix G
Washback Questionnaire

Appendices

WASHBACK QUESTIONNAIRE

Dear teachers, you are kindly requested to put a tick (√) on the statement that best describes your opinion.

N	Statement	Strongly Agree	Agree	Strongly Disagree	Disagree	Not Sure
1	I always analyze my students' needs before I design my oral course.					
2	I develop the teaching method with which I feel more comfortable.					
3	I follow the curriculum and the established syllabus.					
4	I see no importance to teach a new topic that will not be examined.					
5	My main objective is to design adequate courses that develop learners' general oral proficiency.					
6	I teach learners to prepare them be good language speakers.					
7	Examinations do not oblige me to teach selected topics.					
8	I emphasize on teaching speaking rather than on scoring.					
9	I do not design my tests according to what I have taught in the classroom					
10	I do not test only what I taught, I keep learners expect everything to develop their proficiency.					

Thank You Very Much Indeed



Appendix I
Oral proficiency Test

Appendices



ORAL PROFICIENCY TEST

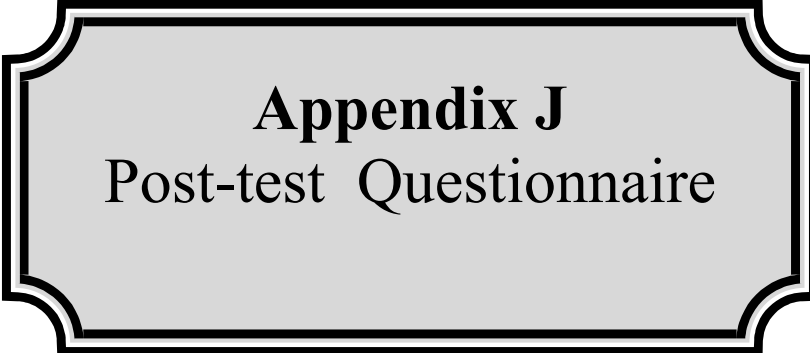
English Department

Second year Oral Proficiency Test
Module : Oral Production

LMD2 G1
Semestre 2

Lab.1.

Stage	Description	Duration	Purpose
I	Warm-up	Less than two minutes	Greetings questions, Exchanging of everyday social amenities;
II	Level-check	Five minutes	Tell us about a perfect vacation you dream about. What stands between you and complete happiness? If you had the opportunity to travel another planet, would you go?
III	Probes	Three minutes	Thought-provoking questions Questions designed to spur learners' oral proficiency. If this phase is successful then this is a good indicator of a good oral proficiency.
IV	Wind-Down	Less than two minutes	Thanking the interviewees.



Appendix J
Post-test Questionnaire

Post-test Questionnaire

Instructions:

- ✓ There are no wrong or right answers.
- ✓ Read each statement carefully. If the statement reflects your experience in taking a test, please place a ✓ in the appropriate box strongly agree, agree; strongly disagree, disagree.
- ✓ Be totally honest with yourself.

Item	Statement	Strongly Agree	Agree	Strongly Disagree	Disagree	Not sure
8.	The oral course was designed to help me get ready for the test					
9.	Examinations influence my learning process.					
10.	Examinations oblige me to revise certain topics.					
11.	The current oral course help me improve my oral proficiency.					
12.	Learning test-taking techniques is most important in class.					
13.	My performance on the test reflects my oral abilities.					
14.	Examinations influence my learning process every year.					
15.	I got worried Immediately before taking the test.					
16.	In oral exams I focus on m performance rather than scores.					
17.	I prefer having my oral test					

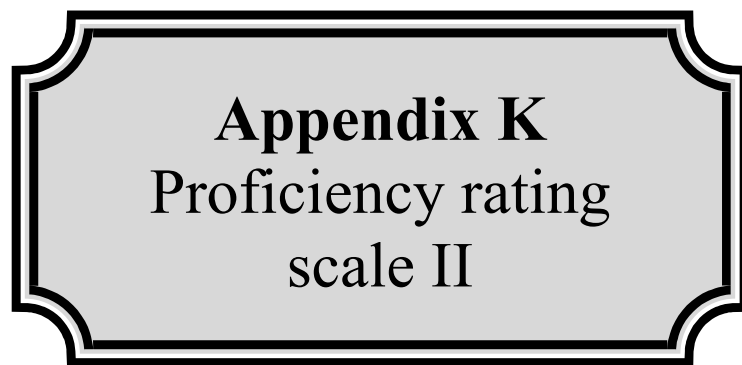
Appendices

	using the computer to record my answers.					
18.	I'll start feeling uneasy just before getting my test score.					
19.	While taking the test I had an uneasy, upset feeling.					
20.	I seem to defeat myself while working on tests like speaking test.					
21.	During the test I found myself thinking about the consequences of failing					
22.	I froze up on the test.					
23.	During the test I got so nervous that I forgot facts I really knew.					
24.	The harder I worked at taking the test, the more confused I got.					
25.	During the test, I got perplexed that I lost vocabulary words I knew					
26.	I am able to handle successfully different communicative skills.					
27.	Oral exam, I prefer one-to-one interview with the teacher					
28.	Thoughts of doing poorly interfered with my concentration on the test.					
29.1	After the test was over I tried to stop worrying about it, but I just couldn't.					
30.	I felt unsure and tense while					

Appendices

	taking the Oral test.					
31.	I wish examinations like oral did not bother me so much					
32.	I prefer talking circle tests rather than face-to-face interview with the teacher					

Thank you very much for your collaboration...



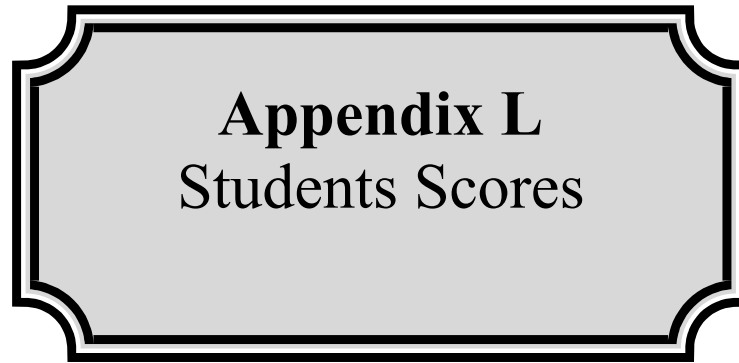
Appendix K
Proficiency rating
scale II

Appendices

Proficiency rating scale II

	Statement	All the time	Most of the time	Occasionally	Rarely	Not At all
1	Comprehension: Demonstrates poor and little Mastery of the topic					
2	Fluency : Control of pace is poor					
3	Vocabulary: Ineffective oral communication due grammar mistakes, incoherence					
4	Pronunciation: Uses unclear pronunciation					
5	Accuracy: Makes inaccurate definitions of concepts					
6	Grammar: Speech full of grammar mistakes					
7	show frequent long pauses, repetition of teacher's words,					
8	Oral production consists of isolated words and perhaps a few high-frequency phrases					
9	able to satisfy the necessities of a basic communication retrieving learned utterances					
10	Able to handle successfully most uncomplicated communicative tasks and social situations					

Criteria	Description	1	2	3	4	5	Descripti on
Accent	Foreign like					→	Native like
Grammar	Inaccurate						Accurate
Vocabulary	Inadequate						Adequate
Fluency	Uneven						Even
Comprehension	Incomplete						Complete

A decorative rectangular box with a double-line border and rounded corners. The text inside is centered and reads "Appendix L" on the top line and "Students Scores" on the bottom line.

Appendix L
Students Scores

Appendices

Tests Scores

Students	Test -1- HR	Test -1- AR	Test -1- R	Test 2 HR	Test 2 AR	Test 2 R
1.	12	09	11	12	10	13
2.	14	11	12	13	12	12
3.	13	12	10	14	11	15
4.	12	10	13	13	10	13
5.	11	12	12	14	10	15
6.	10	11	08	15	12	14
7.	12	08	10	16	13	16
8.	08	04	06	12	11	12
9.	11	08	07	14	12	15
10.	11,5	10	08	13,5	11	14
11.	13	10	10	13	13	12
12.	14	12	10	13,5	12	14
13.	16	13	14	12	10	13
14.	13,5	11	10	14	12	15
15.	12	09	12	13	11	13
16.	11	11	10	12,5	11	15
17.	12,5	10	10	14	12	12
18.	10	07	07	13,5	11	11
19.	10,5	06,5	08	11,5	10	14
20.	12	09	09	13,5	14	15
21.	14	11	11	12,5	13	13
22.	15	12	12	14	11	15
23.	15,5	12,5	13	16	14	16
24.	13,5	10	10	15	11	14
25.	12,5	09	10	14,5	14	14
26.	11	10	07	13	12	14
27.	11,5	11	07	12	11	15
28.	12,5	10	12	11	13	13
29.	11	08	09	15,5	14	16
30.	14	13	13,5	14	14	14
31.	10	11	11,5	12	11	11,5
32.	12	13	12,5	11	12	11,5
33.	12	10	11	14	14,5	14,75
34.	11	11	11	12	09	10
35.	11	10	10,5	13	14	13,5



Appendix M
SPSS Tables

Hypothesis Test Summary

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The categories defined by S1 = Strongly Agree and Agree occur with probabilities 0,5 and 0,5.	One-Sample Binomial Test	,453 ¹	Retain the null hypothesis.
2	The categories defined by S2 = Strongly Agree and Agree occur with probabilities 0,5 and 0,5.	One-Sample Binomial Test	,453 ¹	Retain the null hypothesis.
3	The categories of S3 occur with equal probabilities.	One-Sample Chi-Square Test	,666	Retain the null hypothesis.
4	The categories of S4 occur with equal probabilities.	One-Sample Chi-Square Test	,277	Retain the null hypothesis.
5	The categories of S5 occur with equal probabilities.	One-Sample Chi-Square Test	,666	Retain the null hypothesis.
6	The categories defined by S6 = Strongly Agree and Agree occur with probabilities 0,5 and 0,5.	One-Sample Binomial Test	1,000 ^d	Retain the null hypothesis.
7	The categories defined by S7 = Strongly Agree and Agree occur with probabilities 0,5 and 0,5.	One-Sample Binomial Test	1,000 ^d	Retain the null hypothesis.
8	The categories defined by S8 = Strongly Agree and Agree occur with probabilities 0,5 and 0,5.	One-Sample Binomial Test	1,000 ^d	Retain the null hypothesis.
9	The categories defined by S9 = Strongly Agree and Agree occur with probabilities 0,5 and 0,5.	One-Sample Binomial Test	,453 ¹	Retain the null hypothesis.
10	The categories of S10 occur with equal probabilities.	One-Sample Chi-Square Test	,683	Retain the null hypothesis.
11	The categories defined by S11 = 5,000 and 3,000 occur with probabilities 0,5 and 0,5.	One-Sample Binomial Test	1,000 ^d	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

¹Exact significance is displayed for this test.

Pre Test Norms SPSS

```

RELIABILITY
/VARIABLES=EXAM course curriculum prescribed slected anxious test totest
abilities affect
/SCALE('ALL VARIABLES') ALL
/MODEL=ALPHA.
    
```

**Reliability
Scale: ALL VARIABLES**

Case Processing Summary

		N	%
Cases	Valid	7	87,5
	Excluded ^a	1	12,5
	Total	8	100,0

a. Listwise deletion based on all variables in the procedure.

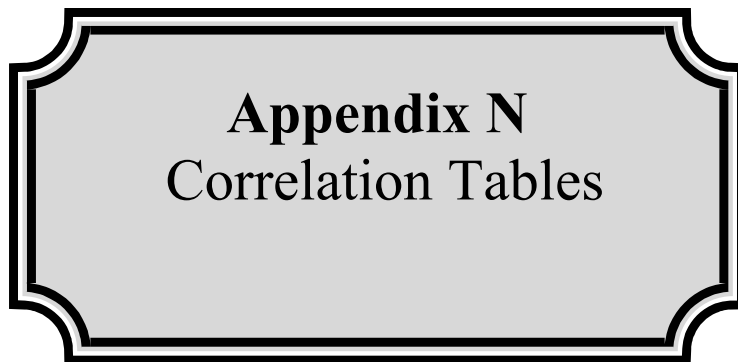
Reliability Statistics

Cronbach's Alpha	N of Items
,961	10

rrelations

Descriptive Statistics

	Mean	Std. Deviation	N
1.	2,8571	1,46385	7
2.	4,4286	,53452	7
3.	3,5714	1,27242	7
4.	2,4286	1,27242	7
5.	3,4286	1,39728	7
6.	2,2857	1,60357	7
7.	3,5714	1,13389	7
8.	3,7143	1,38013	7
9.	3,4286	1,39728	7
10.	4,2857	,48795	7



Appendix N
Correlation Tables

Appendices

Correlation Table

	Correlations						
	EXAM	course	Curriculum	prescribed	slected	anxious	test
Pearson Correlation	1	,730	,767*	,575	,850*	,730	,861*
Sig. (2-tailed)		,062	,044	,177	,015	,062	,013
N	7	7	7	7	7	7	7
Pearson Correlation	,730	1	,560	,665	,606	,806*	,629
Sig. (2-tailed)	,062		,191	,103	,149	,029	,131
N	7	7	7	7	7	7	7
Pearson Correlation	,767*	,560	1	,750	,964**	,642	,891**
Sig. (2-tailed)	,044	,191		,052	,000	,120	,007
N	7	7	7	7	7	7	7
Pearson Correlation	,575	,665	,750	1	,723	,910**	,726
Sig. (2-tailed)	,177	,103	,052		,066	,004	,065
N	7	7	7	7	7	7	7
Pearson Correlation	,850*	,606	,964**	,723	1	,680	,977**
Sig. (2-tailed)	,015	,149	,000	,066		,093	,000
N	7	7	7	7	7	7	7
Pearson Correlation	,730	,806*	,642	,910**	,680	1	,720
Sig. (2-tailed)	,062	,029	,120	,004	,093		,068
N	7	7	7	7	7	7	7
Pearson Correlation	,861*	,629	,891**	,726	,977**	,720	1
Sig. (2-tailed)	,013	,131	,007	,065	,000	,068	
N	7	7	7	7	7	7	7
Pearson Correlation	,801*	,645	,963**	,746	,938**	,721	,867*
Sig. (2-tailed)	,030	,117	,001	,054	,002	,068	,011
N	7	7	7	7	7	7	7
Pearson Correlation	,850*	,606	,964**	,723	1,000**	,680	,977**
Sig. (2-tailed)	,015	,149	,000	,066	,000	,093	,000
N	7	7	7	7	7	7	7
Pearson Correlation	,533	,730	,499	,844*	,524	,943**	,559
Sig. (2-tailed)	,218	,062	,255	,017	,228	,001	,192
N	7	7	7	7	7	7	7

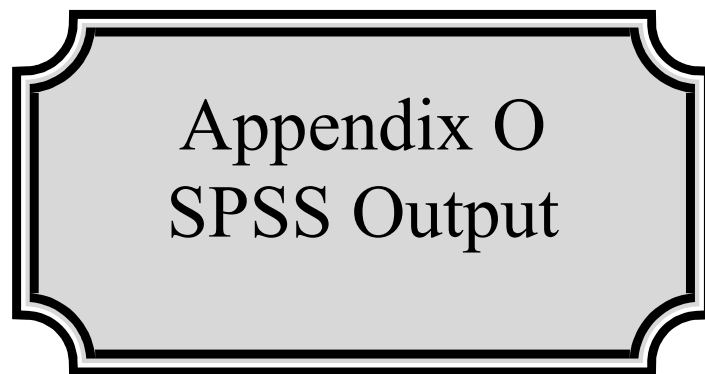
*. Correlation is significant at the 0.05 level (2-tailed).

** . Correlation is significant at the 0.01 level (2-tailed).

NONPAR CORR

/VARIABLES=EXAM course curriculum prescribed slected anxious test totest
abilities affect

/PRINT=KENDALL TWOTAIL NOSIG



Appendix O
SPSS Output

Appendices

Students Test Scores I

	notel1	notel2	note3	Notell1	Notell2	var	var	var	var	var	var	var	var	var	var	var
1	12,00	9,00	11,00	12,00	10,00											
2	14,00	11,00	12,00	13,00	12,00											
3	13,00	12,00	10,00	14,00	11,00											
4	12,00	10,00	13,00	13,00	10,00											
5	11,00	12,00	12,00	14,00	10,00											
6	10,00	11,00	8,00	15,00	12,00											
7	12,00	8,00	10,00	16,00	13,00											
8	8,00	4,00	6,00	12,00	11,00											
9	11,00	8,00	7,00	14,00	12,00											
10	12,00	10,00	8,00	13,50	11,00											
11	13,00	10,00	10,00	13,00	13,00											
12	14,00	12,00	10,00	13,50	12,00											
13	16,00	13,00	14,00	12,00	10,00											
14	13,50	11,00	10,00	14,00	12,00											
15	12,00	9,00	12,00	13,00	11,00											
16	11,00	11,00	10,00	12,50	11,00											
17	12,50	10,00	10,00	14,00	12,00											
18	10,00	7,00	7,00	13,50	11,00											
19	10,00	6,50	8,00	11,50	10,00											
20	12,00	9,00	9,00	13,50	14,00											
21	14,00	11,00	11,00	12,50	13,00											
22	15,00	12,00	12,00	14,00	11,00											
23	15,50	12,50	13,00	16,00	14,00											

Students Tests scores II

	comprehension	Pronunciati...	Fluency	Grammar	Vocabulary	var	var	var	var	var	var	var	var	var	var	var
1	3,00	2,00	2,00	4,00	2,00											
2	2,00	2,00	2,00	3,00	3,00											
3	3,00	1,00	1,00	4,00	3,00											
4	2,00	3,00	3,00	2,00	4,00											
5	4,00	1,00	1,00	3,00	2,00											
6	2,00	3,00	2,00	4,00	3,00											
7	4,00	3,00	3,00	2,00	3,00											
8	4,00	2,00	3,00	2,00	1,00											
9	2,00	1,00	2,00	3,00	4,00											
10	2,00	3,00	3,00	2,00	4,00											
11	3,00	1,00	2,00	2,00	3,00											
12	2,00	3,00	2,00	2,00	2,00											
13	3,00	1,00	2,00	3,00	2,00											
14	4,00	2,00	2,00	3,00	2,00											
15	3,00	2,00	2,00	1,00	3,00											
16	3,00	1,00	1,00	2,00	3,00											
17	1,00	2,00	2,00	3,00	3,00											
18	2,00	3,00	1,00	3,00	1,00											
19	2,00	3,00	2,00	4,00	4,00											
20	2,00	2,00	2,00	3,00	4,00											
21	2,00	2,00	1,00	4,00	3,00											
22	3,00	2,00	2,00	4,00	3,00											

Appendices

Washback Statistical Data CYCLE I

	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
1	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
2	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
3	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
4	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
5	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
6	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	4,00	5,00	5,00	5,00	4,00	4,00	5,00
7	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	4,00	5,00	4,00	4,00	4,00	4,00	5,00
8	5,00	5,00	5,00	5,00	5,00	5,00	4,00	5,00	4,00	4,00	4,00	5,00	4,00	4,00	5,00
9	5,00	5,00	5,00	5,00	5,00	5,00	4,00	5,00	4,00	4,00	4,00	4,00	4,00	4,00	5,00
10	5,00	5,00	5,00	5,00	5,00	5,00	4,00	5,00	4,00	4,00	4,00	4,00	4,00	4,00	5,00
11	5,00	5,00	5,00	5,00	5,00	4,00	4,00	5,00	4,00	4,00	4,00	4,00	3,00	4,00	5,00
12	5,00	5,00	5,00	5,00	5,00	4,00	4,00	5,00	4,00	4,00	4,00	4,00	3,00	4,00	5,00
13	5,00	4,00	5,00	5,00	5,00	4,00	4,00	5,00	4,00	4,00	4,00	4,00	3,00	4,00	4,00
14	5,00	4,00	5,00	5,00	4,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00	3,00	4,00	4,00
15	4,00	4,00	4,00	5,00	4,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00	3,00	3,00	4,00
16	4,00	4,00	4,00	5,00	4,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00	3,00	3,00	4,00
17	4,00	4,00	4,00	5,00	4,00	4,00	4,00	4,00	3,00	4,00	2,00	4,00	3,00	3,00	4,00
18	4,00	4,00	4,00	5,00	4,00	3,00	4,00	4,00	3,00	4,00	2,00	4,00	3,00	3,00	4,00
19	4,00	4,00	4,00	5,00	4,00	3,00	4,00	4,00	3,00	3,00	2,00	3,00	3,00	2,00	4,00
20	4,00	4,00	4,00	5,00	2,00	3,00	3,00	4,00	3,00	3,00	2,00	2,00	2,00	2,00	4,00
21	4,00	4,00	4,00	5,00	2,00	3,00	2,00	4,00	3,00	3,00	2,00	2,00	2,00	2,00	4,00
22	4,00	4,00	4,00	4,00	2,00	3,00	2,00	4,00	3,00	2,00	2,00	2,00	2,00	2,00	4,00
23	4,00	4,00	4,00	4,00	1,00	3,00	2,00	4,00	3,00	2,00	2,00	2,00	2,00	2,00	3,00

Washback Stristical Data Cycle II


	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25
1	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
2	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
3	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
4	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00
5	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	5,00	4,00	5,00	5,00	5,00
6	4,00	5,00	5,00	5,00	4,00	4,00	5,00	4,00	5,00	5,00	4,00	5,00	4,00	5,00	5,00
7	4,00	5,00	4,00	5,00	4,00	4,00	5,00	4,00	5,00	5,00	4,00	5,00	4,00	4,00	5,00
8	4,00	4,00	4,00	5,00	4,00	4,00	5,00	4,00	5,00	5,00	4,00	4,00	4,00	4,00	5,00
9	4,00	4,00	4,00	4,00	4,00	4,00	5,00	4,00	5,00	5,00	4,00	4,00	4,00	4,00	5,00
10	4,00	4,00	4,00	4,00	4,00	4,00	5,00	4,00	5,00	5,00	3,00	4,00	4,00	4,00	5,00
11	4,00	4,00	4,00	4,00	3,00	4,00	5,00	4,00	5,00	4,00	3,00	4,00	4,00	4,00	5,00
12	4,00	4,00	4,00	4,00	3,00	4,00	5,00	4,00	5,00	4,00	3,00	4,00	4,00	4,00	5,00
13	4,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00	5,00	4,00	5,00	3,00	3,00	4,00	5,00
14	3,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00	4,00	4,00	2,00	3,00	4,00	3,00	4,00
15	3,00	4,00	4,00	4,00	3,00	3,00	4,00	3,00	4,00	4,00	2,00	3,00	3,00	3,00	4,00
16	3,00	4,00	4,00	4,00	3,00	3,00	4,00	3,00	4,00	4,00	2,00	3,00	3,00	3,00	4,00
17	3,00	4,00	2,00	4,00	3,00	3,00	4,00	3,00	4,00	4,00	2,00	3,00	3,00	3,00	4,00
18	3,00	4,00	2,00	4,00	3,00	3,00	4,00	3,00	4,00	4,00	2,00	3,00	3,00	3,00	4,00
19	3,00	3,00	2,00	3,00	3,00	2,00	4,00	2,00	4,00	4,00	2,00	2,00	3,00	3,00	4,00
20	3,00	3,00	2,00	2,00	2,00	2,00	4,00	2,00	3,00	4,00	4,00	2,00	2,00	2,00	4,00
21	3,00	3,00	2,00	2,00	2,00	2,00	4,00	2,00	3,00	4,00	4,00	2,00	2,00	2,00	3,00
22	3,00	2,00	2,00	2,00	2,00	2,00	4,00	2,00	3,00	4,00	3,00	2,00	2,00	2,00	3,00
23	3,00	2,00	2,00	2,00	2,00	2,00	3,00	2,00	2,00	4,00	2,00	2,00	2,00	2,00	3,00



Appendix P
SPSS Certificate

SPSS Certificate





Appendix Q
Speechace Invoice

Appendices



speechace

SpeechAce LLC

2133, 5th Avenue, 406, Seattle, WA, USA 98121

Contact: abhishek@speechace.com

QUOTE

Invoice Number: N/A

Date: February 24, 2020

Client: University Tlemcen / Contact: Houda Djebbari

Item	Quantity	Rate	Amount
Speechace IELTSAce (www.ielsace.com) bulk subscription 300 annual users	1	\$1,725	\$1,725

Subtotal: USD 1,725.00

WA Sales Tax (9.6%): USD 0

Total: USD 1,725.00