



Sudan University of Science and Technology

College of Graduate Studies

College of Education

**Washback-based Approach of the Achievement Test in
Assessing the Learning Process**

الاختبار التحصيلي المبني على طريقة الانعكاس واثره على العملية التعليمية

**A thesis Submitted in Fulfillment of the Requirements for
PhD Degree in English Language**

Prepared by: Ibrahim Ahmed Eisa Abdallah

Supervised by: Dr. Ishraga Bashir Elhassan

Co-supervised by Dr: Einas Ahmed Abdel Rahman

2019



Dedication

To my parents, wife, daughters and friends who were behind every step of achievement throughout the study.

Acknowledgements

First of all, I would like to convey my gratitude to the Lord, God Almighty, for the strength He gave during the course of the study.

I am deeply indebted to the skilled and patient guidance of my Supervisor Dr. Ishraqa Bashir Elhassan and my Co- Supervisor Dr. Enas Ahmed Abdulrahman for their great academic support and valuable guidance.

I am extremely grateful to my friends and colleagues Dr. Sabir Mirgani Ali Ahmed, Dr. Fatherahman Abdulwahid Gubara and Dr. Ahmed Khadir Osman for their encouragement throughout the stages of this study. I would also like to take the opportunity to thank the participating institutions, seminars participants, teachers and students who cooperatively made this study possible.

Abstract

This study investigates the influence of the achievement test results on developing the EFL learning process. The study adopts experimental and descriptive methods to analyze and verify the results in which frequencies, percentages, means and their standard deviation were used. The analysis was carried out through the Statistical Package for Social Sciences (SPSS) technique. The study sample comprises (30) students of first-year secondary school in the eastern region of Saudi Arabia who have been exposed to the test and (30) teachers of EFL in Saudi Arabia who responded to the questionnaire. Two tools have been used for collecting data relevant to the study, namely (pre/post) test to the students and a questionnaire to the teachers of English as a foreign language. These tools were judged by arbitrators and provided answers to the research questions and hypotheses. The researcher designed a specific teaching plan aiming at improving the learners' language knowledge. For carrying out the plan, all available techniques, teaching materials and aids were exploited during the plan. The results have shown great improvements in the performance of learners and the study has come up with findings that revealed the impact of test on assessing the EFL learning process. The study offered some recommendations and suggestions for further studies that will encourage EFL learners to be aware of the positive and negative influence of test results on developing the learning process.

(Arabic Version)

المستخلص بالعربي

هذه الدراسة تبحث في أثر انعكاس نتائج الاختبار التحصيلي في تطوير عملية التعلم لدى دارسي اللغة الإنجليزية كلغة اجنبية. اعتمدت الدراسة المنهج التجريبي والوصفي لتحليل النتائج وتأكيد النسب المؤية والوسيط والانحراف المعياري. تم استخدام التقنية الإحصائية (SPSS) لتحليل البيانات المتعلقة بالدراسة. وقد تم اختيار عينة الدراسة من (30) دارسا من طلاب الصف الاول الثانوي بالمنطقة الشرقية في المملكة العربية السعودية للإجابة على اسئلة الاختبار وعينة اخري بحجم (30) معلما من معلمي اللغة الانجليزية في المملكة العربية السعودية لاداء الاستبيان. تم استخدام اثنين من الأدوات لجمع البيانات ذات الصلة بموضوع البحث والتي تتمثل في اختبار قبلي(تشخيصي) واختبار بعدي(تحصيلي) لطلاب السنة الأولى ثانوي واستبيان لمعلمي اللغة الانجليزية كلغة اجنبية للإجابة على اسئلة البحث وفرضياته. تم تحكيم ادوات البحث بواسطة عدد من المختصين في المجال. قام الباحث بتصميم خطة دراسية لفترة اربع اشهر زمنية لرفع مستوى الطلاب في مادة اللغة الانجليزية. وتم الاستعانة بكل الوسائل التعليمية والادوات المتاحة من اجل رفع مستوى الطلاب المعرفي في اللغة. وقد اظهرت النتائج تحسنا واضحا في أداء الدارسين. وعليه قدمت الدراسة بعض المقترحات والتي تتمثل في الاستفادة من اثر نتائج الاختبار التحصيلي في تطوير عملية التعلم لدى دارسي اللغة الانجليزية كلغة اجنبية. و قدمت الدراسة بعض التوصيات والمقترحات التي تشجع دارسي اللغة الإنجليزية في ادراك الاثر السلبي والايجابي لنتائج الاختبار التحصيلي ودوره في تطوير عملية التعلم.

Table of Contents

Items	Pages
Dedication	i
Acknowledgement	ii
Abstract	iii
Abstract (Arabic version)	iv
Table of Contents	v
List of Tables	xi
List of Figures and Diagrams	xvii
Definition of Terms	xx
CHAPTER ONE	
INTRODUCTION	
1.0 Introduction	1
1.1 Statement of the problem	3
1.2 Objectives of the study	3
1.3 Questions of the study	3
1.4 Hypotheses of the study	4
1.5 Significance of the study	4
1.6 Scope of the study	4

1.7 Methodology of the study	5
1.8 10 Summary	5
CHAPTER TWO	
LITERATURE REVIEW AND PREVIOUS STUDIES	
2.0 Introduction	6
Part one: Literature Review	6
2.1 Theoretical Frame work of Washback	
2.2 The Origin of Washback	8
2.3 Definitions of Washback	9
2.4 Types of Washback	13
2.4.1 Positive Washback	15
2.4.2 Negative Washback	17
2.4.3 Pedagogical Implications of Washback	21
2.5 The Importance of Studying Washback	22
2.6 Relation of Testing to Teaching and learning	25
2.7 Language Testing and Washback Effect	28
2.8 Functions of Washback	29
2.9 Washback as Criterion for Developing and Evaluating Language Tests	32
2.10 Mechanism of Washback	33

2.10.1 Models of Washback	35
2.10.1.1 Hughes’s Washback Model	36
2.10.1.2 Bailey’s Washback Model	37
2.10.1.3 Burrow’s Washback Models	39
2.11 Washback a phenomenon Leading to Educational Innovation	41
2.12 Areas Affected by Washback	42
2.12.1 Washback on Syllabuses and Curriculumms	45
2.12.2 Washback and Teaching to the Test	46
2.12.3 Washback on Teaching Methodology	47
2.12.4 Washback on Teacher Factors	48
2.12.5 Washback on Language Learning	52
2.12.6 Washback on Test Takers	53
2.12.7 Washback on Materials	54
2.12.8 Washback on Lesson Contents	56
2.12.9 Washback on Learning Outcomes	57
2.12.10 Washback Impact on Society and Education System	58
2.12.11 Washback Impact on Feelings and Attitudes	59
2.12.12 Washback on Resources	59
2.12.13 Washback on the School	59

2.12.14 Washback on the Exam	60
2.13 Testing the Different Language Areas	60
2.13.1 Types of Tests	61
2.13.2 Language Testing	64
2.13.3 The Language Areas to be Tested	66
2.13.4 Testing the Language Areas	67
2.13.5 Making a Test Format	74
2.13.6 Giving a Test	75
2.13.7 The Characteristics of a Good Test	76
2.14 Part Two: Previous Studies	78
Introduction	78
Comments on the Previous Studies	80
2.15 Summary	88
CHAPTER THREE METHODOLOGY	
3.0. Introduction	90
3.1 Methods of Analysis (Statistics)	90
3.2. Population of the Study	90
3.3. Sample of the Study	90

3.4 Tools of Data Collection	91
3.5 Test Design and Format	91
3.5. 1 Test Procedures	92
3.5.2 Test Administration	92
3.6 Teaching Plan (Treatment)	93
3.7 The Statistical Treatment of the Test	93
3.7.1 Test Validity	93
3.7.2 Test Reliability	94
3.8 Questionnaire Design and Format	94
3.8.1 Questionnaire procedures	96
3.8.2 Questionnaire Administration	97
3.9 The Statistical Treatment of the Questionnaire	97
3.9.1 Questionnaire Validity	97
3.9.2 Questionnaire reliability	98
3.10 Summary	98
CHAPTER FUOR	
DATA ANALYSIS AND DISCUSSION	
4.0 Introduction	99
4.1 The Responses to the Students' Pre- Post Tests	99
4.2 Analysis and Discussion of Students' Test	100

4.3 The Responses to the Teachers' Questionnaire	103
4.4 Analysis and Discussion of Teachers' Questionnaire	103
4.5. Testing the Study Hypotheses	133
4.5.1 Results of the First Hypothesis:	133
4.5.2 Results of the Second Hypothesis:	136
4.5 Summary	142
CHAPTER FIVE	
MAIN FINDINGS, CONCLUSIONS, RECOMMENDATIONS AND SUGGESTIONS FOR FURTHER STUDIES	
5.0. Introduction	143
5.1 Main Findings	143
5.2 Conclusion	144
5.3 Recommendations	145
5.4 Suggestions for Further Studies	146

List of Tables

No	Title	Page
2.1	Table (2.1) Summary of Positive washback Hughes (1989)	19
2.2	Table (2.2) Summary of Negative washback Hughes (1989)	19
2.3	Table (2.3) Hughes's trichotomy of backwash model (1993)	36
3.1	Components and items of students' pre and post tests	92
3.2	Components and items of teacher questionnaire	96
4.1	The Subjects' Results in the Pre-and Post-Tests	99
4.2	T-test for the differences between the scores in pre and posttests in reading part	100
4.3	T-test for the differences between the scores in pre and posttests in grammar part	101
4.4	T-test for the differences between the scores in pre and posttests in vocabulary part	101
4.5	T-test for the differences between the scores in pre and posttests in conversation part	101
4.6	T-test for the differences between the scores in pre and posttests in writing part	102
4.7	T-test for the differences between the scores in pre and posttests for overall part	102
4.8	The Frequency Distribution for the Respondent's Answers of Statement No (1)(Curriculum and Syllabus)	103

4.9	The Frequency Distribution for the Respondent's Answers of Statement No (2) A	104
4.10	The Frequency Distribution for the Respondent's Answers of Statement No (3) A	104
1.11	The Frequency Distribution for the Respondent's Answers of Statement No (4) A	105
4.12	The Frequency Distribution for the Respondent's Answers of Statement No (5) A	106
4.13	The Frequency Distribution for the Respondent's Answers of Statement No (6) A	106
4.14	The Frequency Distribution for the Respondent's Answers of Statement No (7) A	107
4.15	The Frequency Distribution for the Respondent's Answers of Statement No (1)B (Textbooks and Materials)	108
4.16	The Frequency Distribution for the Respondent's Answers of Statement No (2) B	108
4.17	The Frequency Distribution for the Respondent's Answers of Statement No (3) B	109
4.18	The Frequency Distribution for the Respondent's Answers of Statement No (4) B	110
4.19	The Frequency Distribution for the Respondent's Answers of Statement No (5) B	110
4.20	The Frequency Distribution for the Respondent's Answers of	111

	Statement No (6) B	
4.21	The Frequency Distribution for the Respondent's Answers of Statement No (7) B	112
4.22	The Frequency Distribution for the Respondent's Answers of Statement No (8) B	112
4.23	The Frequency Distribution for the Respondent's Answers of Statement No (9) B	113
4.24	The Frequency Distribution for the Respondent's Answers of Statement No (10) B	114
4.25	The Frequency Distribution for the Respondent's Answers of Statement No (1)C (Teaching Methods)	114
4.26	The Frequency Distribution for the Respondent's Answers of Statement No (2) C	115
4.27	The Frequency Distribution for the Respondent's Answers of Statement No (3) C	116
4.28	The Frequency Distribution for the Respondent's Answers of Statement No (4) C	116
4.29	The Frequency Distribution for the Respondent's Answers of Statement No (5) C	117
4.30	The Frequency Distribution for the Respondent's Answers of Statement No (6) C	118
4.31	The Frequency Distribution for the Respondent's Answers of Statement No (7) C	118
4.32	The Frequency Distribution for the Respondents' Answers of Statement No (8) C	119

4.33	The Frequency Distribution for the Respondent's Answers of Statement No (9) C	120
4.34	The Frequency Distribution for the Respondent's Answers of Statement No (1) D (Classroom Tasks and Activities)	120
4.35	The Frequency Distribution for the Respondent's Answers of Statement No (2) D	121
4.36	The Frequency Distribution for the Respondent's Answers of Statement No (3) D	122
4.37	The Frequency Distribution for the Respondent's Answers of Statement No (4) D	122
4.38	The Frequency Distribution for the Respondent's Answers of Statement No (5) D	123
4.39	The Frequency Distribution for the Respondent's Answers of Statement No (6) D	124
4.40	The Frequency Distribution for the Respondent's Answers of Statement No (1) E (EFL Skills and Elements)	124
4.41	The Frequency Distribution for the Respondent's Answers of Statement No (2) E	125
4.42	The Frequency Distribution for the Respondent's Answers of Statement No (3) E	126
4.43	The Frequency Distribution for the Respondent's Answers of Statement No (4) E	126
4.44	The Frequency Distribution for the Respondent's Answers of	127

	Statement No (5) E	
4.45	The Frequency Distribution for the Respondent's Answers of Statement No (1)F(Beliefs, Attitudes, and Perception Related to the Test and Teaching)	128
4.46	The Frequency Distribution for the Respondent's Answers of Statement No (2) F	128
4.47	The Frequency Distribution for the Respondent's Answers of Statement No (3) F	129
4.48	The Frequency Distribution for the Respondent's Answers of Statement No (4) F	130
4.49	The Frequency Distribution for the Respondent's Answers of Statement No (5) F	130
4.50	The Frequency Distribution for the Respondent's Answers of Statement No (6) F	131
4.51	The Frequency Distribution for the Respondent's Answers of Statement No (7) F	132
4.52	The Frequency Distribution for the Respondent's Answers of Statement No (8) F	132
4.53	Table No (4.53) testing the first hypothesis of the study	134
4.54	Table No (4.54) testing the second hypothesis of the study	136
4.55	Table No (4.55) Explains the study hypotheses of Teaching Methods	139
4.56	Table No (4.57) Explains the study hypotheses of EFL Skills and Elements	140

4.57	Table No (4.57) Explains the study hypotheses of EFL Skills and Elements	140
4.58	Table No (4.58) Explains the study hypotheses of Beliefs, Attitudes, and Perception Related to the Test and Teaching	141

List of figures

No	Title	Page
2.1	Bailey's washback model (1996)	38
2.2	Burrow's washback models (1998)	40
4.1	T-test for the differences between the scores in pre and posttests for overall.	102
4.2	Teachers are aware of the objectives of the syllabus and curriculum.	103
4.3	The present syllabus and curriculum can enhance EFL learning process	104
4.4	Teachers teach every section in the textbook (<i>Traveller</i>) although some sections are unlikely to be tested in the examination.	105
4.5	Teachers do not care about the syllabus and curriculum while teaching their students.	105
4.6	Teachers feel pressure to cover the syllabus before the final test.	106
4.7	The test tests the overall competence of students in English (listening, speaking, reading, and writing).	107
4.8	Teachers give more attention to teaching to the syllabus opposed to practicing the test items.	107
4.9	Teachers follow and communicate the lesson objectives with the students while planning their lessons.	108
4.10	The textbook (<i>Traveller</i>) covers exercises that the curriculum has claimed	109
4.11	Teachers think their students do not seriously study the textbook materials.	109
4.12	Teachers skip certain topics in the textbook because they are less likely to be tested in the examination.	110
4.13	The textbook (<i>Traveller</i>) is well-suited to developing the students' communicative competence	111
4.14	Teachers rely on test papers and sample questions to prepare their students for the final test.	111

4.15	Teachers find interest in teaching textbook materials.	112
4.16	If teachers teach the whole textbook (<i>Traveller</i>), their students will perform badly in the test	113
4.17	Teachers use authentic materials along with the textbook for the students' practice of English language in the class.	113
4.18	Teachers use modern aids and equipment to teach and practice English language skills.	114
4.19	Teachers teach in a way that their students understand everything.	115
4.20	Teachers teach and explain the text in English.	115
4.21	Teachers encourage their students to ask questions during the class.	116
4.22	Teachers encourage their students to speak English in the class.	117
4.23	Teachers use Arabic along with English to make their students understand better.	117
4.24	Teachers teach whatever they think important to teach, no matter whether it is important or not for the test.	118
4.25	The role of teachers of English is to transmit knowledge to their students through explaining texts and giving examples.	121
4.26	Teachers do not make their students practice how to learn and speak English language but make them practice on how to answer questions in the test.	119
4.27	Teachers believe that the test score in English is an appropriate indicator of a student's English ability.	120
4.28	Teachers ignore the task and activities that are not directly related to passing the test.	121
4.29	Teachers give model tests to the students to do better in the final test.	121
4.30	Teachers spend more time teaching grammar because they think grammar is more likely to be tested.	122
4.31	Teachers make their students practice and solve the questions of the past tests.	123

4.32	The test discourages the teachers to teach English language.	123
4.33	Teachers teach test-taking strategies, especially when the test date gets closer.	124
3.34	The test influences the teacher's decision on which language skills are more important to be taught.	125
4.35	Listening is practiced in the class.	125
4.36	Speaking is practiced in the class.	126
4.37	Reading is practiced in the class.	127
4.38	Writing is practiced in the class.	127
4.39	Teachers feel pressure from their authority to improve their students' test score.	128
4.40	Students can score good marks without improving their English language proficiency.	129
4.41	Teachers get feedback on their teaching from the test results.	129
4.42	Students suffer from anxiety and tension for the test.	130
4.43	The present testing system helps the students improve language proficiency.	131
4.44	Teachers could teach English better if there were no pressures for good results in the test.	131
4.45	Tests influence students' future career.	132
4.46	Teachers feel embarrassed if their students fail or perform badly in the test.	133

Definition of Terms

A number of key terms are defined as follows in order to establish a consistent and common meaning for them as they are used in this thesis.

1. Achievement Test: It is a test for program evaluation, directly drawn from the content of instruction, and show if students have learnt what has been taught. Henning (1987, p.6).

2. Applied Linguistics: The study of the relationship between theory and practice. The main emphasis is usually on language teaching, but can also be applied to translation, lexicology, among others Richard (2001).

3. Diagnostic test: Diagnostic test is a type of formative evaluation that attempts to diagnose students' strengths and weaknesses vis-a-Vis the course materials. Students receive no grades on diagnostic instruments.

4. Evaluation and Assessment: Evaluation is described as an overall but regular judgment and analysis of teaching, learning, as well as curriculum through systematic collection of data. Assessment looks at the individual language learners, but evaluation checks the whole language-learning program. In assessment data is collected by concentrating on students' moment-by-moment performance in the classrooms "emanating from alternative activities"

Genesee (2001, p.149) while evaluation involves the gathering of data by focusing on teaching performance and learning outcomes .

5. Likert Scale: It is a semantic differential scale that requires subject's to respond to the statements by using a numerical indication of the strength of their feeling towards the object or position described in the statement.

6. Linguistics: It is the scientific study of natural language. Linguistics encompasses a number of sub-fields such as phonology, morphology, semantics and syntax.

7. Micro level impacts: By micro level impacts, they refer to the stake-holders - students and teachers- who are directly affected by the testing procedure and the results (classroom settings).

8. Macro level impacts: They refer to the effects of tests on educational contexts and society in a broader sense (educational and societal system).

9. Standardized test: A standardized test measures language ability against a norm or standard. It is a test that is constructed in accord with detailed specifications, one for which the items are selected after tryout for appropriateness in difficulty and discriminating power, one which is accompanied by a manual giving definite directions for uniform administration and scoring, and one which is provided with relevant and dependable norms for score interpretation.

10. Washback: the influence of language testing on teaching and learning (Cheng, Watanabe, Curtis, 2004). Therefore, it is a type of an impact, which relates to the effects of high-stakes tests on classroom practices –particularly teaching and learning. How assessment instruments affect educational practices and beliefs Cohen (1994, p. 41).

11. Validity: According to some definitions “Validity is that quality of the tools used or procedure that enables to measure what it is supposed to measure” Validity of a tool lies mainly in the procedure adopted for constructing it.

12. Reliability: It refers to the reliability of any test, to obtaining the same results if the same measurement is used more than one time under the same conditions. Also, it is defined as the degree of the accuracy of the data that the test measures.

CHAPTER ONE

INTRODUCTION

1.0 Background of the Study

English as a foreign Language (EFL) tests have played a significant socio-economic role in modern societies. They are used to make many crucial decisions about people by policy makers who are therefore able to exert power and the authority in the context in which tests are used. Language tests can also be used to communicate the educational priorities to the stakeholders in the results of the tests. Language testing is a complex area which continues to create debate between language researchers and test developers. However, while applied linguists discuss the exact nature of language proficiency and how it can be assessed, educators still need to make decisions about the language ability of students as accurately and efficiently as possible.

Testing is often seen as both a necessary evil and a vehicle for affecting educational change. Test scores are used by policymakers in education as tools to control admission, promotion, placement and graduation. Moreover, the importance of English Examinations cannot be denied: it is a criterion for admission into higher education, it supports the processes of self-evaluation and improvement of educational institutions, and it serves as instrument and basis both for research work and for granting certain educational benefits. So, the English tests can be seen as an initial step towards a long-term goal and raising the standards of communicative competence of EFL learners.

In recent years, there has been growing interest among testers in the field of education, in the effects, both desirable and undesirable, of tests and the concepts of ‘test impact’ and ‘test washback’. Impact is defined by Davies (1999: 79) as “the effect of a test on individuals, on educational systems and on society in general.” The term washback, or ‘backwash’ as it is sometimes referred to, can be broadly defined

as the effect of testing on teaching and learning, and is therefore a form of impact. It is then a concept which includes several specialist areas in the field of applied linguistics.

The impact of tests or examination system on the learning process has been widely discussed in the field of education Morris (1990). Due to his wide use, tests, particularly high-stakes tests, may exert an influence on teachers and students with an associated impact on what happens in classrooms. The phenomenon that tests influence what and how teachers teach and students learn is described as the washback effect (Alderson & Wall, 1995). Hughs (2003) defined washback as “a part of the impact a test may have on learners and teachers, on educational systems in general, and on society at large” (p. 53). For decades now, scholars have debated the potential impact of tests. Although some scholars have also considered that tests have the potential to bring about a favorable impact on learning process (e.g., Bailey, 1996; Wall, 2000) review article noted that most researchers believe that tests induce deleterious effects. However, in a country where the debate on the importance of learning English as a second language is far from over, the issue of test impact or washback-based approach of the test in learning process gets overlooked.

In order to answer research questions, I first provide an overview of theoretical and practical issues in washback studies and the educational context, to give a clear idea of the context of the study. Therefore, Chapter One includes the introduction and the base on which this study is conducted. Chapter Two is devoted to the presentation of theoretical framework of the study; it defines the field of the study and literature review on washback-based approach to testing and the previous studies. Chapter Three covers the methods adopted in collecting data besides, showing how the problem of the research is tackled. Chapter Four deals with analysis of data and their

interpretation. The main findings, recommendations, suggestion for further studies and conclusion are summarized in Chapter Five.

1.1 Statement of the Problem

The researcher as a teacher of English language has been met by a countless number of complaints and discontent among teachers of English language as well as students about the results achieved by students in their final exams. Many students who performed well in Exams and obtained full marks, unfortunately were unable to use the language communicatively or properly in context of life and manage their learning process. Therefore, this phenomenon would hypothetically be associated with a negative washback of the test, which is the effect of test results on the syllabus, the language materials, the teacher and the learner, and hence hinders the learning process including the basic skills of EFL– listening, speaking, reading and writing negatively. The researcher would also like to discover, whether the EFL achievement tests at Secondary Schools at their current situation have positive or negative washback, using schools as a case study.

1.2 Objectives of the Study

The study aims to:

1. investigate the impact of washback-based approach of the achievement test on developing the EFL learning process at Qatif Secondary School.
2. explain how washback-based approach of the achievement test can promote EFL learners' proficiency.

1.3 Questions of the Study

1. What is the impact of the washback-based approach of the achievement test on developing the EFL learning process?
2. How can the washback-based approach of the achievement test promote EFL learners' proficiency?

1.4 Hypotheses of the Study

1. Washback- based approach of the achievement test has positive impact on developing the EFL learning process.
2. Washback-based approach of the achievement test can promote EFL learners' proficiency.

1.5 Significance of the Study

The importance of investigating more about washback-based approach of the achievement test is crucial. Firstly, this study provides some of the needed knowledge and understanding to help teachers and language test developers to design instruction and tests that properly account for washback. Secondly, language classroom teacher has a vested interest in knowing how washback affects the learning process, and how to best recognize the effects of washback-based approach, either positive or negative, on her/his teaching. Thirdly, learning more about washback in this environment has important implications for teacher education programs designed to train new instructors, to introduce valuable classroom techniques and strategies, and to review important issues in the development of course syllabi and teaching materials. It is hoped that this study will be of great value to teachers of English language and their students when dealing with investigation into the washback-based approached effects on EFL achievement tests. It is also hoped that the findings of this study might be useful to teachers of English as a foreign language in general.

1.6 Scope the Study

The study is limited to the:

- 1. limits of Subject:** The subject of the study was about Washback-based Approach of the Achievement Test: Assessing the Learning Process.
- 2. limits of Place:** The study was applied and carried out at First Secondary School in the Eastern Region of Saudi Arabia as a case study.
- 3. limits of Time:** The study was carried out and applied during the academic year: 2016/2017).

1.7 Methodology of the Study

The study adopted the descriptive and experimental methods throughout the research. Two tools were used that aiming at gathering the needed data. A diagnostic and achievement (pre / post) as primary tools to reveal the positive or negative impact of washback-based of the test that encounters first year secondary students in learning process. A questionnaire was distributed to EFL teachers in checking their points of views in terms of this issue.

1.8 Summary

This introductory chapter was concerned with presentation of background of the study, statement of the problem, objectives of the study, questions of the study, hypotheses of the study, significance of the study, scope of the study, methodology of the study and finally the definition of key terms.

CHAPTER TWO

LITERATURE REVIEW AND PREVIOUS STUDIES

2.0 Introduction:

Judging from the volume of research in recent years, the washback-based approach effect is still to be adequately defined and analyzed. While there is consensus that washback incorporates the effects of tests on teaching and learning, researches have not agreed on what washback is, what it might look like, or how it works. There have only been a limited number of washback studies, and invariably, researchers call for further investigations that would establish what washback is and even whether it exists.

This chapter contextualizes the current study within the framework of a critical review of the relevant literature on the theoretical framework of washback, origin of washback, definitions of washback, its connection to impact, positive and negative connotations, importance of studying washback, language testing and washback effect, the function and mechanism of washback, areas that affected by washback. In addition to, models of achievement test washback and it presents an overview of some major washback studies done in different parts of the world. It also summarizes the related researches with emphasis on the washback effects and the impact of the achievement test on learning process with particular attention on testing the different language areas. Some pervious researches and studies in the field of this study have been stated by the researcher at the end of this chapter.

Part One: Literature Review

2.1 Theoretical Framework of Washback

Washback, sometimes referred to as backwash Biggs (1994) can be generally understood as the effect of an examination on teaching and learning. Hughes (2003). He restricted the use of the term ‘washback’ to:

“classroom behaviors of teachers and learners rather than the nature of printed and other pedagogic material” (p. 118). They would also consider washback to be what teachers and learners do that “they would not necessarily otherwise do” (p. 117).

Wall (1997) makes a clear distinction between washback and test impact.

The latter would refer to the effect of a test on “individuals, policies or practices, within the classroom, the school, the educational system or society as a whole” (cited in Cheng and Curtis, 2004, p.4). Other researchers do not make that distinction and they consider that narrow and wider effects can be included under the term washback. For the purposes of this study, washback will be understood in the wider sense that is, including what some scholars call ‘impact’.

The concept of washback has been associated with validity. Morrow (1986) refers to “washback validity” to describe the quality of the relationship between testing and teaching and learning”. For Messick (1996) washback is part of construct validity, and it is an inherent quality of any kind of assessment, especially when the results are used for important decisions. For him, washback contributes to the consequential aspect of construct validity, but information about the operative level of test validity should help one distinguish test washback from the effects of good or bad educational practices regardless of the quality of the test. So, washback in itself is not a reliable criterion to establish test validity. It is other test properties, like authenticity and directness that are likely to produce washback.

Messick (1996) claims that:

“If a test is deficient because it has construct under representation or construct-irrelevant characteristics, then good teaching cannot be considered an effect of the test, and conversely, if a test is construct-validated, but there is poor teaching, then negative washback cannot be associated with the test. Only valid test can increase the likelihood of positive washback”

2.2 The Origin of Washback

Although the subject of the effects of examinations has long been discussed in the literature of general education (Vernon, 1956); it has been common in the literature on testing that washback, as it is known for us now, has come to attract the attention of test researchers only at the beginning of the 1990's. Before that date, testing specialists and applied linguists used different terms to refer to the idea of examination influence. Language testing researchers have realized that the emergence of the concept washback is the result of considerable reforms and advances that have taken part in the domain of language testing mainly during the last two decades at the end of the twentieth century.

In a comprehensive study on how the concept washback has come to emerge, Tsagari (2007) designs an artificial time framework divided into three distinct but successive phases that clearly display how the concept has evolved overtime in the scene of language testing. These phases are the "pre-1990's", the "1990's", and the "post 1990's". Tsagari identifies that the initial phase was mainly characterized as the period of time when writers recognized the existence of the examination influence phenomenon but no one accounted for it. Second, the 1990's phase was thought to be different from the previous one and was basically dominated by the publication of a seminal paper by two prominent language testing researchers, Alderson and Wall, who are greatly indebted the fact they were the first who questioned the nature of examination influence; and more importantly, they managed to re-conceptualize this phenomenon by proposing a set of hypotheses. The third phase, the "post 1990's", or as Tsagari names it the 'reality phase', was significant since substantial models of washback have been developed in order to accurately explain and analyze the nature of this phenomenon. Such a way of looking at the subject is seen by testing experts to be a serious attempt and a step forward in the study of washback in language testing.

2.3 Definitions of Washback

It is widely accepted in language teaching that major tests and examinations inevitably have an influence both within the education system and beyond it. There are two terms which are used to refer to this influence: impact and washback. The broader concept is impact, which Wall (1997:291) defined as "... any of the effects that a test may have on individuals, policies, or practices, within the classroom, the school, the educational system or society as a whole." Bachman and Palmer (1996) note the complex nature of the phenomenon and suggested that impact should be considered within the social context. Taking into account variables such as social goals and values, the educational system the test is used in, and the potential outcomes of its use. Washback is a form of impact and the term is commonly used to refer specifically to the effects of a test on teaching and learning.

In general, the term 'washback' refers to any influence a test may have on teachers and learners, either positive or negative, and either at the micro or macro level. At the micro level, the test affects students and instructors; at the macro level, it affects the instructional system of the institution. Pierce (1992: 687) specifies classroom pedagogy, curriculum development, and educational policy as the areas where washback has an effect. On the other hand, Alderson and Hamp-Lyo (1996) took a view of washback which concentrated more on the effect of the test on teaching. They referred to washback as "... the influence that writers of language testing, syllabus design and language teaching believe a test will have on the teaching that precedes it" (ibid: 280).

Concern about the influence of assessment is not confined to language teaching. A phrase that is often used in this context is 'teaching to the test'. For example, Swain (1985) commented that:

"It has frequently been noted that teachers will teach to a test: that is, if they know the content of a test and/or the format of a test, they will teach their students accordingly" (p. 43). This is often perceived as an unacceptable practice in education in general.

Mehrens (1991) presents the following guidelines concerning appropriate instructional strategies:

1. a teacher should not engage in instruction that attenuates the ability to infer from the test score to the domain of knowledge/skill/ability of interest;
2. it is appropriate to teach the content domain to which the user wishes to infer;
3. it is appropriate to teach test-taking skills;
4. it is inappropriate to limit content instruction to a particular test item format;
5. it is inappropriate to teach only objectives from the domain that are sampled on the test;
6. it is inappropriate to use an instructional guide that reviews the questions of the latest issue of the test;
7. it is inappropriate to limit instruction to the actual test questions;
8. it is appropriate to teach toward test objectives if the test objective comprise the domain objectives;
9. it is appropriate to ensure that students understand the test vocabulary; and
10. one cannot teach only the specific task of a performance assessment.

Here the discussion extends to how certain kinds of test preparation might 'pollute' test scores and make them less reliable. In terms of washback, Mehrens indicates that points 2, 3, 8 and 9 reflect practices which would have a positive effect, while the remainder can be seen as indications of negative washback.

This distinction between positive and negative effects is also recognized by language testers. We can expect students and teachers to prepare in some way for test, therefore, some forms of test preparation must be accepted as appropriate activities. Davies et al (2001: 210) describe this 'test wise ness':

Familiarity, or lack of it, with item or task types, the kinds of text used, or appropriate techniques (such as reading the questions relation to a text before reading the text itself or apportioning a suitable time to each section of the test), are likely to affect a test taker's score.

This type of preparation that familiarizes students with the requirements of the test and supports them practicing on test-like materials. If the test in question is one that is in keeping with current theories of language, we might expect it is more likely to encourage positive washback. Thus, if the skills that are to appear on the exam correlate with the objectives of the curriculum then the washback of that exam could be said to be positive. But, on the other hand, if the instruction is adjusted to correspond with the exam's subject matter, then this has a negative effect on the curriculum and the exam can be seen as having negative washback.

Alderson and Hamp-Lyons summarize some typical concerns regarding negative washback to the curriculum (1996: 281, original citations included):

1. Narrowing of the curriculum (Madaus, 1988; Cooley, 1991)
2. Lost instructional time Smith et al (1989)
3. Reduced emphasis on skills that require complex thinking or problem solving (Fredericksen, 1984; Darling-Hammond and Wise, 1985)
4. Test score 'pollution' or increases in test scores without an accompanying rise in ability in the construct being tested.

Some washback has only negative connotations. For example, Spolsky (1995: 55) defined washback as a "term better applied only to accidental side-effects of examinations, and not to those effects intended when the first purpose of the examination is control of the curriculum" and spoke of the "...inevitable outcome in narrowing the educational process..." (ibid.). He uses vocabulary tests to illustrate what he calls the 'crux of the washback problem'. While vocabulary tests may be a

quick measure of language proficiency, once they are established as the only form of assessment, the backwash to instruction resulted in the tests becoming a measure of vocabulary learning rather than language proficiency. Spolsky suspects that “a good proportion of the Asian students who took the test, were more highly motivated to pass the examination than to learn English”, a fact that he attributes in part to the “tendency to cramming” among many Asian students (1995: 64). For Davies et al (1999:225), negative washback occurs when the test items are based on an outdated view of language which bears little relationship to the teaching curriculum (ibid.). The example they give is a useful one: “If...the skill of writing is only tested by multiple choice items, then there is great pressure to practice such items rather than to practice the skill of writing (ibid.)” A test which is structure based and prevents proficiency teaching becoming communicative in nature, can be said to have a negative washback effect.

According to Alderson (2001) test washback is commonly seen as a negative form of impact as it forces teachers to do things they do not really want to do. Similarly, Wall and Alderson (1993) reasoned that if the aims, activities, or marking criteria of the textbook and the exam contain no conflicts and the teachers accept and work towards these goals, then this is a form of positive washback. Negative washback would be evidenced in the exam having a distorting or restraining influence on what is being taught and how. Alderson and Banerjee (2001) acknowledge that tests have the potential to be ‘levers for change’ in education if one accepts the argument that if bad tests have a negative impact then it should be possible for a good test to have good washback.

Testing has not always been portrayed as having a negative effect on education and educational systems. An early advocate in favor of working for positive washback in language testing was Hughes (1989), who argued that to promote beneficial washback one should "base achievement tests on objectives rather than on detailed

teaching and textbook content and they will provide a truer picture of what has actually been achieved. He also saw it as crucial that learners and teachers knew and understood the test because no matter how good the potential washback a test might have, if the requirements were not understood, the potential was unlikely to be realized. More generally, Bailey stated that “A test will generate positive washback to the learner if the test-taker buys into the assessment process”. She goes on to comment that “A test will promote beneficial washback to the extent that it is based on sound theoretical principles which are widely accepted in the field...” (1996: 176), and that "A test will yield positive washback to the learner and to the program to the extent that it utilizes authentic tasks and authentic texts." (p. 276). Bailey concluded, "a test will promote beneficial washback to programs to the extent that it measures what programs intend to teach."

To sum up the matter of defining the concept, Bailey (1996) stated that washback is positive or negative depending on whether or not the test promoted the learners' language development (as opposed to their perceived progress or to their test-taking skills). The literature on the topic implies that the higher the stakes involved with proficiency tests, the more likely those assessments are going to affect the preparation in the course curriculum prior to that assessment, as teachers will tend to “teach to the test” and students will tend to focus more, if not exclusively, on the skills or subjects they expect to meet in the exam.

2.4 Types of Washback

Alderson and Wall (1993) distance themselves from a simplistic assumption about the way a test can influence behaviors. Therefore, they developed washback hypotheses according to what is influenced: teaching, learning, content, rate, sequence, degree, depth, attitudes and also the number of teachers or learners affected by a test. Which hypotheses will be put forward depends on the nature of the test, the educational context, and the nature of the decisions that are taken on the basis of the

test results. Actually, there seems to be a number of variables in society, education, and schools that determine how washback will appear.

When studying washback, it is also possible to focus on participants (teachers, students, material developers, publishers), process (actions by participants towards learning), and products (what is learned and the quality of learning), as suggested in Hughes's trichotomy model (Hughes, 1993 as cited in Bailey, 1996). Watanabe (2004) proposes disentangling the complexity of washback by conceptualizing it in terms of: Dimension (specificity, intensity, length, intentionality and value of the washback), aspects of learning and teaching that may be influenced by the examination, and the factors mediating the process of washback being generated (test factors, prestige factors, personal factors, macro-context-factors).

Usually researchers focus on one aspect or type of washback. In Alderson and Wall's study in Sri Lanka (Alderson & Wall, 1993; Wall, 1996), the introduction of a test of English as a foreign language proved to produce faster changes in the content of teaching than changes in teaching methodology. Cheng (1997), in the preliminary results of a study of the washback effect of the Hong Kong Certificate of Education Examination in English in Hong Kong secondary schools, reports that washback effect "works quickly and efficiently in bringing about changes in teaching materials [...] and slowly and reluctantly and with difficulties in the methodology teachers employ" (p.1). Cheng introduces the term 'washback intensity' to refer to the "degree of washback effect in an area or a number of areas that an examination affects most" (p.7).

Generally, washback can be analyzed according to two major types: positive and negative, depending on whether it has a beneficial or harmful impact on educational practices (Hughes, 1989). This section explores positive and negative washback in terms of both the classroom setting and the educational/political system.

Alderson and Wall (1993, as cited in Fulcher and Davidson 2007) include different factors in their washback hypotheses:

“If teachers use tests to get their students to pay more attention to lessons and to prepare more thoroughly, it is positive washback. If teachers fear poor results and the associated guilt which might lead to the desire for their students to achieve high scores in tests, it might be a reason for teaching to the test. Consequentially, teachers narrow the curriculum and produce negative washback”.

2.4.1 Positive Washback

Some researchers, on the other hand, strongly believe that it is feasible and desirable to bring about beneficial changes in language teaching by changing examinations so-called 'positive washback'. This term is directly related to 'measurement-driven instruction' in general education and refers to tests that influence teaching and learning beneficially. In this sense, teachers and learners have a positive attitude toward the test and work willingly toward its objectives. Pearson (1988:107) argued that 'good tests will be more or less directly usable as teaching-learning activities. Similarly, good teaching-learning tasks will be more or less directly usable for testing purposes, even though practical or financial constraints limit the possibilities'. Considering the complexity of teaching and learning, such claim sounds ideal, but rather simplistic. Davies (1985) maintains the view that a good test should be 'an obedient servant of teaching; and this is especially true in the case of achievement testing'. He (1985:8) further argues that 'creative and innovative testing... can, quite successfully, attract to itself a syllabus change or a new syllabus which effectively makes it into an achievement test.' In this case, the test no longer needs to be only an 'obedient servant': rather it can also be a 'leader'.

However, there are rather conflicting reactions toward whether there is positive or negative washback on teaching and learning. Some linguists argued that paid coaching classes, which were intended for preparing students for exams, were not a good use of the time, because students were practicing exam techniques rather than language learning activities. However, Heyneman (1987:262) commented that many proponents of academic achievement testing view 'coach ability' not as a drawback, but rather as a virtue. Pearson (1988:101) looks at the washback effect of a test from the point of view of its potential negative and positive influences on teaching. According to him,

“a test's washback effect will be negative if it fails to reflect the learning principles, and/or course objectives to which they supposedly relate, and it will be positive if the effects are beneficial and 'encourage the whole range of desired changes”.

Alderson and Wall (1993:118-117), on the other hand, stress that the quality of the washback effect might be independent of the quality of a test. Any test, good or bad, can be said to have beneficial or detrimental washback. Whatever changes educators would like to bring about in teaching and learning by whatever assessment methods, it is worthwhile to investigate first the broad educational context in which an assessment is introduced since other forces exist within the society, education, and schools that might prevent washback from appearing (Alderson and Wall, 1993: 116). Heyneman (1987:262) concluded that 'testing is a profession, but it is highly susceptible to political interference. To a large extent, the quality of tests relies on the ability of a test agency to pursue professional ends autonomously'. If the consequences of a particular test for teaching and learning are to be evaluated, the educational context in which the test takes place needs to be investigated. Whether the washback effect is positive or negative will largely depend on how it works and within which educational contexts.

A. Class room setting : Teachers and learners will be motivated to fulfil their teaching and learning goals. Good tests can be utilized and designed as beneficial teaching-learning activities so as to encourage a positive teaching-learning process (Pearson, 1988:107). A creative and innovative test can quite advantageously result in a syllabus alteration or a new syllabus Davis (1985).

B. Educational/ sociétal system : Decisional makers use the authority power of high-stakes testing to achieve the goals of teaching and learning, such as the introduction of new textbooks and new curricula (Shohamy, 1992 ; Wall & Alderson 1993 ; Cheng ; 2005). Tests are encouraged to promote the idea of lifelong learning and encourage people to learn English (Language Testing and Training Centre, 2008).

2.4.2 Negative Washback

Language tests and tests in general are often criticized for their negative influence on teaching so-called 'negative washback. Vernon (1956:166) commented that teachers tended to ignore subjects and activities, which did not contribute directly to passing the exam, and claimed that examinations 'distort the curriculum'. Davies (1968a:125, 1968b), for example, indicates that all too often the washback effect has been bad; designed as testing devices, examinations have become teaching devices; work is directed to what are in effect if not in fact past examination papers and consequently becomes narrow and uninspired. Alderson and Wall (1993:5) refer to 'negative washback' as the negative or undesirable effect on teaching and learning of a particular and, by inference if not direct statement, 'poor' test. In this case, 'poor' usually means 'something that the teacher or learner does not wish to teach or learn.' The tests may well fail to reflect the learning principles and/or the course objectives to which they are supposedly related.

Fish (1988) discovered that teachers reacted negatively to pressure created by public displays of classroom scores, and also found that relatively inexperienced teachers

felt greater anxiety and accountability pressure than did experienced teachers. Noble and Smith (1991a:3) also pointed out that high-stakes testing affected teachers directly and negatively, and that 'teaching test-taking skills and choice worksheets is likely to boost the scores but unlikely to promote general understanding'(1991b:6). Smith (1991b:8) concluded from an extensive qualitative study of the role of external testing in elementary schools that 'testing programs substantially reduce the time available for instruction, narrow curricular offerings and modes of instruction, and potentially reduce the capacities of teachers to teach content and to use methods and materials that are incompatible with standardized testing formats'.

A. Class room setting : The test will lead to the narrowing of content in the curriculum. What students have learned is test language, instead of total phases of understanding Shohamy (1992). Teachers tend to ignore subjects and activities that are not directly related to passing the exam, and tests accordingly alter the curriculum in a negative way Vernon (1956).

Educators experienced negative reactions to the stress brought about by public displays of classroom scores. Inexperienced teachers felt a greater degree of anxiety and pressure for accountability than did teachers with more experience. "Testing programs substantially reduce the time available for instruction, narrow curricular offerings and modes of instruction, and potentially reduce the capacities of teachers to teach content and to use methods and materials that are incompatible with standardized testing formats" Smith (1991).

An increasing number of paid coaching classes are set up to prepare students for exams, but what students learn are test-taking skills rather than language learning activities. Measurement-driven instruction will definitely result in cramming, narrowing the curriculum, focus of attention on those skills that are most relevant to testing, placement of constraints on teachers' and students' creativity and spontaneity, and disparage the professional judgment of educators Madaus (1988).

B. Educational /societal system

Decision makers overwhelmingly use tests to promote their political agendas and to seize influence and control of educational systems Shohamy (1996). Tests are used as a “lever” for change.

To present a clear view of positive and negative washback at both micro-level (classroom settings) and at macro-level (educational and societal system), Tables 1 and 2 is presented below for classification.

Table (2.1) Summary of Positive washback Hughes (1989)

Positive Washback	
Classroom settings	<ol style="list-style-type: none">1. Tests induce teachers to cover their subjects more thoroughly, making them complete their syllabi within the prescribed time limits.2. Tests motivate students to work harder to have a sense of accomplishment and thus enhance learning.3. Good tests can be utilized and designed as beneficial teaching-learning activities so as to encourage positive teaching-learning processes.
Educational/societal system	Decision makers use the authority power of high-stakes testing to achieve the goals of teaching and learning, such as the introduction of new textbooks and new curricula.

Table (2.2) Summary of Negative washback Hughes (1989)

Negative Washback

Classroom settings	<ol style="list-style-type: none"> 1. Tests encourage teachers to narrow the curriculum and lose instructional time, leading to “teaching to the test.” 2. Tests bring anxiety both to teachers and students and distort their performance. 3. Students may not be able to learn real-life knowledge, but instead learn discrete points of knowledge that are tested. 4. Cramming will lead students to have a negative positive toward tests and accordingly alter their learning motivation.
Educational/societal system	Decision makers overwhelmingly use tests to promote their political agendas and to seize influence and control of educational systems.

To summarize, in terms of the classroom setting at a micro level, the positive washback integrates meaningful and innovative learning activities in teachers’ educational methodologies, and thus educators will devote more attention to students’ intentions, interests, and choices. Students at the same time will be encouraged and motivated to work harder. On the other hand, the negative washback is that teachers will usually teach to the test, narrow the curriculum and only focus on what will be tested. Moreover, cramming will be the washback brought by measurement-driven tests, even though there is an ongoing debate as to whether cramming is positive or negative washback. In terms of educational setting, the positive washback is that the authority can use the test to attain its goal of teaching and learning. However, the negative washback is that the authority uses that goal to control and obtain the power of the academic system that will usually place undue pressure and anxiety on school staffs, teachers and even students. In other words, the washback on the side of the educational setting is one coin with two sides, depending on the stakeholder’s point of view.

2.4.3. Pedagogical Implications of Washback

By analyzing the possible positive and negative washback that tests might bring about at micro and macro levels, it seems that teachers play an important role in fostering different types of washback. In other words, the beliefs of the teachers are a critical factor in determining the washback effect. For example, a test, on one hand, will encourage some teachers to think it's important to plan their curricula carefully to meet the needs of the test, but on the other hand, other teachers may think that tests force them to teach what they don't find suitable or appropriate for students. Spratt (2005) has stated that the teacher plays a significant role in determining the types and intensity of washback, and thus, teachers have become the sources of promoting positive washback. Chapman and Snyder J. (2000:462) have expressed a similar view by stating that "it is not the examination itself that influences teachers' behavior, but teachers' beliefs about those changes". As Watanbee (2004) suggested, teachers should be provided with in-service training and be familiar with a wide range of teaching methods.

Tests sometimes are used by schools or school administrations as a "lever" to introduce the innovation of new curricula, but it may change the format of what teachers instruct, not foster an in-depth change of teaching methodologies as a whole. As Wall (2005:283) stated, "examinations cannot influence teachers to change their practices if they are not committed to the new ideas and if they do not have the skills that will enable them to experiment with, evaluate and make appropriate adjustments to new methods". In other words, teachers themselves must conduct the changes and teachers need to have the necessary skills to adapt the changes. Again, teachers play a very crucial role in promoting positive washback or hindering negative washback.

To conclude, there are two major perspectives that teachers should bear in mind. If we are the ones who make the tests, we should try to make a match between what is tested and what is taught by using more direct testing, making sure the test is known by students. Tests are one factor that will lead the teacher to "teach to the test", and

what students learn might be discrete points of language, not the communicative part of language they need in real life. To remedy this, it is desirable to use authentic and direct tests Bailey (1996). If we are responsible for helping students pass the test, we should try our best to learn more teaching methodologies by taking more training courses, engaging in peer observations and utilizing the tests to enhance students' learning while at the same time not inhibiting students' motivation by cramming too much.

As teachers, “we may have limited power to influence high-stakes national and international examinations, but we do have tremendous power to lead students to learn, to teach them language and how to work with tests and test results.” Bailey (2005) All in all, it's the teacher who has the most power to turn it into positive or negative washback.

2.5 The Importance of Studying Washback

The strong influence of high-stakes tests on learning processes has long been accepted in the field of education. In the field of applied linguistics ‘the concept of a test influencing learning in the language learning classroom was rarely discussed until the early 1990s (Andrews, 2004; Bailey, 1996 ‘Cheng, 1997; Elder & Wigglesworth, 1996; Wall, 2000). The term washback became used in the field to refer to the power that high-stakes tests could have on language teaching and learning, although impact or consequences are more commonly used in the field of education. While the concept of washback was earlier only asserted based on anecdotal evidence Burrows (2004), the pioneer evidence-based washback research was carried out by Alderson and Wall (1993). They investigated the effects of the introduction of new tests in Sri Lanka on the teaching of English as a foreign language by secondary school teachers. They found, however, that teachers' lessons remained teacher-centered over the period of two years and students still had little chance to use English in a practical way – although language learning activities and

the design of classroom tests were influenced by the new textbooks .They concluded that the effects of the implementation of new tests were much more limited than expected and that the mechanism of washback was not as straightforward as previously thought .

When studying washback, it is also possible to focus on participants (teachers, students, material developers, publishers), process (actions by participants towards learning), and products (what is learned and the quality of learning), as suggested in Hughes's trichotomy model (Hughes, 1993 as cited in Bailey, 1996) .Watanabe (2004) proposes disentangling the complexity of washback by conceptualizing it in terms of: Dimension (specificity, intensity, length 'intentionality and value of the washback), aspects of learning and teaching that maybe influenced by the examination, and the factors mediating the process of washback being generated (test factors, prestige factors, personal factors, macro-context factors). Usually researchers focus on one aspect or type of washback. In Alderson and Wall's study in Sri Lanka (Alderson & Wall, 1993; Wall, 1996), the introduction of a test of English as a foreign language proved to produce faster changes in the content of teaching than changes in teaching methodology. Cheng (1997) in the preliminary results of a study of the washback effect of the Hong Kong Certificate of Education Examination in English in Hong Kong secondary schools, reports that washback effect "works quickly and efficiently in bringing about changes in teaching materials [...] and slowly and reluctantly and with difficulties in the methodology teachers employ" (p.1). Cheng introduces the term 'washback intensity' to refer to the "degree of washback effect in an area or a number of areas that an examination affects most" (p.7) .

Andrews et al. (2002) finds out in their study that the impact of a test can be immediate or delayed. According to these researchers, washback seems to be associated primarily with 'high-stakes' tests, that is:

“tests used for making important decisions that affect different sectors., for example, determining who receives admission into further education or employment opportunities (Chapman and Snyder, 2000). identify as ‘high’ such situations when admission, promotion, placement and graduation are dependent on the test”.

Cheng (2000) reports on how tests are often introduced into the education system to improve teaching and learning, especially in countries where tests are considered an efficient tool for introducing changes into an educational system without having to change other educational components. In some countries, these tests can be considered “the engine for implementing educational policy” Cheng (2000, p.6).

In recent years, researchers have been making significant inroads into investigating this phenomenon in different social and educational contexts. As a result, the definition as well as the nature and scope of washback have been extensively discussed, and a number of different perspectives have emerged in language testing and ELT research area. Despite the strong link between testing ‘teaching and learning discussed in the field of education, the assertion that a test influences what teachers and students do in the classroom is often based on anecdotal evidence and did not receive much attention from researchers until the early 1990s in the field of applied linguistics .

There is sufficient evidence indicating that teachers’ perceptions of washback seldom overlap the perceptions of test designers or policy. Furthermore, though a good number of washback studies have been carried out during recent years in different countries, the washback effect is still to be adequately defined and analyzed .

2.6 Relation of Testing to Teaching and Learning

There is an in-depth relation between testing, teaching and learning. Test objectives determine the teaching objectives. Testing strongly influences the classroom activities. Tests are assumed to be powerful determiners of what happens in classrooms; and it is commonly claimed that tests affect teaching and learning activities both directly and indirectly.

As mentioned earlier, washback, a term commonly used in applied linguistics, refers to the influence of language testing on teaching and learning. The influence of a test on the classroom is, of course, very important; washback effect can be either beneficial or harmful. Teachers as well as their students tailor their classroom activities to the demands of the test, especially when the test is very important for the future of the students. A high-stakes test is a type of test whose results are seen-rightly or wrongly by students, teachers, administrators, parents, or the general public as the basis upon which important decisions that immediately and directly affect the students are made.

A test can be considered as high-stakes if the test results are perceived by stakeholders (e.g., teachers, students, parents and schools) to have serious consequences, such as graduation, comparison or placement of students, the evaluation of teachers or schools, and/or the allocation of resources to schools (Madaus (1988). High-stakes tests can be norm or criterion-referenced, and internal and external in origin. They offer future academic and employment opportunities based upon the results. The achievement test, the subject of the study, is such a high-stake test. It is given to the students at the end of their school year of education. Students either proceed to further grades after passing the examination or stay in the same grade. Washback is the power of examinations over what takes place in the classroom (Alderson and Wall, 1993, p.115). Numerous explanations of the term

'washback' can be found throughout the published research and literature on language testing. One of the most common definitions sees the concept referred to as the influence of testing on teaching and learning (e.g. Alderson & Wall, 1993; Gates 1995; Cheng & Curtis 2004). They define washback as "the connection between testing and learning" (p.298). Gates (1995) explains washback simply as "the influence of testing on teaching and learning" (p.101). Messick (1996) refers to washback as "... the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning" (p. 241). Pierce (1992) states that the washback effect is sometimes referred to as the systemic validity of a test. Bachman and Palmer (1996, p.29-35) have discussed washback as a subset of a test's impact on society, educational systems, and individuals. Alderson and Wall (1993) consider washback as the way that tests are perceived to influence classroom practices, and syllabus and curriculum planning. Cohen (1994) describes washback in terms of how assessment instruments influence educational practices and beliefs. Achievement examinations are often used as instruments to select students as well as a means to control a school system and are commonly believed to have an impact on teaching and learning.

Given that achievement test has exerted an influence on teachers and students with an associated impact on what happens in classrooms, such a phenomenon is denoted as "washback" or "backwash" (Alderson, 1986; Morrow, 1986; Pearson, 1988; Hughes, 1989; Morris, 1990). A test is an experience that the teacher creates to serve as a basis for grading a learner in order to group them according to a laid down standard by a government or an institution. A test is a method that generally requires some performance or activity on the part of either the testee or the tester, or both. There is a set of techniques, procedures, test items that constitutes an instrument of some sort. Such a type of test is commonly believed to have an impact on teaching and learning. Every test does not carry the same weight and importance. High stakes tests influence

the way students and teachers behave, the content and methodology of teaching programs, attitudes towards the value of certain educational objectives and activities, the academic employment options that are open to individuals, and may have significant long-term implications for education systems and for the societies in which they are used. According to Alderson and Wall (1993), the notion that testing influences teaching is referred to as 'backwash' in general education circles, but it has come to be known as 'washback' in applied linguistics (p. 11). Washback and backwash are now interchangeably used in both EFL and ESL research of applied linguistics (Bailey, 1999). Washback or backwash has been defined as "a part of the impact a test may have on learners and teachers, on educational systems in general, and on society at large" Hughes (2003, p. 53). In recent years, there has been growing interest among the testers in the field of education, in the effects, both desirable and undesirable, of tests and the concepts of 'test impact' and 'test washback'. Impact is the consequence of a test on individuals, on educational systems and on society in general. The term 'washback' or 'backwash' as it is sometimes referring to, can be broadly defined as the effect of testing on teaching and learning, and is therefore a form of impact. In language tests, how far the learners have attained language proficiency has to be measured. There is a natural tendency for both teachers and learners to tailor their classroom activities to the demands of the test, especially when the test is very important for the future of the learners, and the pass rates are used as a measure of teachers' success. There is a consensus among the educators that the contents of classroom instruction should be decided on the basis of clearly understood educational goals, and examinations should try to ascertain whether these goals have been achieved. The influence of examinations on second/foreign language (SL/FL) teaching and learning has become an area of significant interest for testers and teachers alike. Negative washback is said to create a narrowing of the curriculum in the classroom so that teachers and learners focus solely on the areas to be tested. On the other hand, there have been attempts to generate positive washback by means

of examination reform to encourage teachers and learners to adopt more modern communicative approaches to language learning. When the examination does that, it forces learners and teachers to concentrate on these goals; and the washback effect on the classroom is very beneficial .

Testing has been used for decades but concerns about its influence have recently increased. Davies et al. (2000) define 'impact' by as "the effect of a test on individuals, on educational systems and on society in general" (P. 79). With this increased concern, the influence of tests has been officially termed as 'washback' or 'backwash' and used as an impact in the field of language testing. Washback appears to be a concern in education in general. This study, however, focuses on washback on SL/FL education. Specifically, the EFL test in the Higher Secondary examination is the subject matter of the present study .

2.7 Language Testing and Washback Effect

Test is a tool of any exam and possesses central position in the framework of an examination system. Focus, form, content and manipulation of a test can affect validity of exam and ultimately affects the quality of English language teaching and learning that aims at imparting language skills. Washback has become a well versed common academic phenomenon in all the institutions. Comprehensively washback can be defined as the influence of testing on teaching and learning Gates (1995).

Washback is not something new it has always been present in the academic career of every student and teacher in the form of the question: Will this be included in the test? Naturally, teachers and students are inclined to adapt teaching and learning according to the requirement of test especially when the test results would have a bearing on their future Buck (1988). Washback has become an important aspect of

present day education because test scores have played a decisive role in the achievement of educational and social pursuits Cheng (1997).

Despite their beneficial or harmful washback, public exam and test have maintained their place because they hold a mirror to all the other activities going on under the umbrella term of teaching-learning process and other processes in education system. It is important to understand that washback is a neutral term that can be considered as positive or negative (Bachman & Palmer, 1996). On the basis of its scope Bachman and Palmer (1996) differentiated test impact from test washback. They considered test impact a ‘macro level’ a phenomenon having social and institutional implications while observed test washback at ‘micro level’ limited to teaching, learning, teachers and students. In the same way, Wall (1997) claimed that test impact is any kind of effect that an exam or test might have on individuals, policies and other educational practices, while test washback is the effect of test specifically on teaching and learning.

Considering distinction between test impact and test washback it was inferred that it is neither easy to measure negative washback effect of exams nor to modify exams to exert a positive effect because washback not only influences teaching and learning but it is also influenced by many other factors.

2.8 Functions of Washback

It has been stated in language testing that tests can serve a number of functions ranging from measuring students' level of competence and knowledge to imposing the effectiveness of teachers and schools. For many educators, a test is often seen as a means by which decision makers usually come to make a judgment on how instruction is carried out and learning is going on, and the extent to which the set-out objectives have been attained. Traditionally, tests used to be at the end of the teaching operation to provide a diagnosis of the effects of teaching and learning. Nevertheless,

with the advances and changes made in the domain of testing, and how the latter is conceived, a test can also be developed to be used at the beginning or in the middle of the teaching/ learning process in order to influence this process and serve specific functions. This view is derived from the realization of test power and its manifestation with regard to examination decisions based on test results for individuals, educational system, and society as a whole Hsu (2009:50). In discussing the functions of language tests through which washback occurs in actual teaching environments, Wall (1993) refers to a number of reviews of those tests and the influence they have on the system they are introduced into. One of the most comprehensive reviews is the one that was introduced by Eckstein and Noah (1993). In its essence, Eckstein and Noah provided a historical account for the functions and influences of some examples of tests as crucial by which they take important decisions for some precise purposes. For the two researchers, the first documented use of written, public examination systems occurred under the Han Dynasty in China, about 200 B.C. The main functions of these examinations were to select candidates for entry into the government services. In other words, the candidates were used to break the monopoly over government jobs enjoyed by the aristocratic Feudal system. Another function was to check patronage and corruption. As an example of this function was Britain where people could gain entry into higher education or the professions of strengths. An important consequence of these examinations was the establishment of numerous public schools, which aimed at preparing students for examination. The third function of examination, as presented by Eckstein and Noah, was to encourage levels of competence and knowledge amongst those who were entering government services or professions. The intention was to design examinations which reflected the demands of the target situation, and students could have to develop skills which were relevant to the work they hoped to get in the future. The fourth function was that of allocating sparse places in higher education. At this level, examinations were used to as a means of selecting the most able candidates for

the available places. This type of examinations is quite the same to what is referred to as the placement tests in the testing literature in the present time. The fifth Washback in Language Testing: a Review of the Concept and its Implications to Innovation in Education function, in this illustration, was to measure and impose the effectiveness of teachers and schools. Eckstein and Noah again used Britain as an example describing how, at a certain time, the government set up a system of examinations through the allocation of considerable funds. The amount of funding that a school received depended on how its students performed. However, this system had serious unintended consequences and at least had failed to achieve the expected objectives. The final function, in Eckstein and Noah set of examples, was limiting curriculum differentiation. In Britain, in the 19th and 20th centuries, there was a remarkable resistance to the idea of centralized education, and all schools had the freedom to decide on their own curriculum and their means of assessment. With the establishment of certificate examinations, the schools had a common target they could aim for, and all these schools turned to teach the curriculum that can help better in doing well in the examinations that are relevant to these certificates (Eckstein, Noah, 1993:5-17). Therefore, this series of functions of tests, exposed above, are typical situations where these tests were used to exert influence- or let us say a washback effect-on the final outcomes to suit the desired intentions of those in authority to make and improve their policies. Shohamy et al. (1996) comments on this assumption and notes that 'the power and authority of tests and external examinations 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' (P.299). In this sense, school wide examinations are used by principals and administrators to enforce learning, which in classrooms, tests and quizzes are used by teachers to impose discipline and motivate learning. Given this status of tests and public examinations, a systematic study of the functions of tests in learning and teaching is essential.

2.9 Washback as a Criterion for Developing and Evaluating Language Tests

Positive washback, by whatever name, has recently been recognized as one of the main criteria for evaluating language tests. In his 1979 book, *Language Tests at School*, Oller identifies the key characteristics of a good test as being reliability, validity, practicality (also called "feasibility"), and instructional value- the last being most closely related to current conceptions of washback.

In 1988 Hughes wrote about the introduction of a needs-based test of English at the university level in Turkey. This article reports on a fascinating case study of test development and implementation in the face of serious resistance. (Specific findings will be discussed below.) Hughes concluded the paper by saying "potential backwash effect should join validity and reliability in the balance against practicality"(ibid., p. 146).

In discussing the development and implementation of a new national exam in Hong Kong, which would include a direct test of speaking ability, Andrews and Fullilove (1994) expressed the test development team's concern that "as far as possible the test should embody the characteristics of a 'good' test. In particular, [the test development group] kept in mind considerations of validity (especially face and content validity), reliability and washback" (p. 64). Indeed, these authors state that the decision to include a costly oral component in this high-stakes exam represented a desire to enhance its validity and improve the positive washback effect it was expected to exert.

Washback has received even more attention as an evaluative criterion recently, with the advent of communicative language testing. For instance, one of Morrow's (1991) five criteria for evaluating communicative language tests is the idea that such tests should "reflect and encourage good classroom practice" (p. 111). In describing a test development project called the *Communicative Use of English as a Foreign Language*, Morrow states: "This [i.e., washback] is a major concern underlying the

design of tests; indeed, in many ways the tests themselves have drawn on 'good' classroom practice in an attempt to disseminate this to other classrooms" (ibid.). Morrow says that a "conscious feedback loop between teaching and testing, in terms of not only content but of approach, is a vital mechanism for educational development" Shohamy (1993b).

The washback concept in language test development has been actively utilized by test developers working in Canada. One influential use of washback in communicative language testing comes from the test development team at the Ontario Institute for Studies in Education, which created the secondary school French test *A Vous La Parole*. One of the team's four tenets for communicative test design was the principle, "Work for washback" - the notion that communicative tests should be explicitly designed to bring about positive washback (Green, 1985, pp. 218-223). Wesche (1987, cited in Pierce, 1992) report that promoting positive washback was also a major concern in the development of the Ontario Test of English as a Second Language.

It is observed that "there has been a recent renewal of interest in the link between good teaching and good testing" (p. xi). They also note that washback, along with validity, reliability and practicality, is now "one of the Big Four considerations in evaluating the worth of a test" (ibid.).

2.10 Mechanism of Washback

Traditionally, tests come at the end of the teaching and learning process for evaluative purposes. However, with the advent of high-stakes achievement test system nowadays, the direction seems to be reversed. Testing usually comes first before the teaching and learning process. When examinations are commonly used as levers for change, new textbooks will be designed to match the purposes of a new

test, and school administrative and management staff, teachers and students will work harder to achieve good scores on the test. In addition, many more changes in teaching and learning can happen as a result of a particular new test. However, the consequences may be independent of the original intentions of the test designers.

Shohamy (1993:2) pointed out that 'the need to include aspects of test use in construct validation originates in the fact that testing is not an isolated event; rather, it is connected to a whole set of variables that interact in the educational process'. Moreover, Messick (1989) recommended a unified validity concept, in which he shows that when an assessment model is designed to make inferences about a certain construct, the inferences drawn from that assessment model should not only derive from test score interpretation but also from other variables in the social context. As early as 1975, Messick (1975:6) pointed out that 'researchers, other educators, and policy makers must work together to develop means of evaluating educational effectiveness that accurately represent a school or district's progress toward a broad range of important educational goals.' In this context, he stated that

“it is incumbent upon the measurement research community to make the case that the introduction of any new high-stakes examination system should include more provisions for paying greater attention to investigations of both the intended and unintended consequences of the system than has been typical of previous test-based reform efforts”.

Exploring the mechanism of such an assessment function, Bailey (1996:262-264) cites Hughes' trichotomy (1993) to illustrate the complex mechanism by which washback works in actual learning context.

2.10.1 Models of Washback

There have been few attempts to describe a model of how a test can influence teaching and learning. This may indicate the difficulty of finding patterns of the way tests influence teachers and students. The impact of an assessment seems to depend not only on the quality of the assessment itself and the way the results are used, but also the context in which the assessment is introduced and administered and the beliefs held by stakeholders such as teachers and students.

During recent years, though a good number of washback studies have been carried out, the washback models are still to be adequately defined and analyzed. In the field of applied linguistics, there seem to have been some attempts to create a model which might illustrate the mechanism of washback. The models of washback discussed below evolve as more research findings became available and a clearer picture of the nature of washback emerged. Thus, the models illustrate the shift in views of washback over the past nearly 20 years. The traditional model of washback emerges in the early 1990s prior to the study by Alderson and Wall (1993). It is characterized as the trichotomy model proposed by Hughes(1993)

Washback models, in general, have been adapted from models or frameworks suggested in language testing, EFL and educational innovation literature. A common characteristic of these washback models is that they tend to highlight what washback looks like and who is affected but do little to address the factors that contribute to the phenomenon. In other words, “process” is less understood than “participants” and “products”. Besides, the products in these models/hypotheses refer mainly to teaching and learning washback, not to the aspects of washback that might impact society. Some specific models that have been proposed in washback literature, and how these they have been developed, are discussed in this section.

2.10.1.1 Hughes's Washback Model

Hughes's (1993) model is a pioneer washback model in applied linguistics. In discussing the complex mechanisms through which washback occurs in actual teaching and learning environments, Hughes (1993) introduces a concept of trichotomy and argues for distinguishing between participants, processes and products in both teaching and learning, recognizing that all three may be affected by the nature of a test. In the Hughes's model (Table 2.3) 'participants' are students, teachers, administrators, materials developers and publishers, whose perceptions and attitudes towards their work may be affected by a test. In his unpublished paper cited by Bailey (1996), and Cheng and Curtis (2004), Hughes (1993) made a distinction between participants, process, and products :

Table (2.3) Hughes's trichotomy of backwash model (1993)

(a)	Participants – students, classroom teachers, administrators, materials developers and publishers, whose perceptions and attitudes toward their work may be affected by a test
(b)	Processes – any actions taken by the participants which contribute to the process of learning
(c)	Products – what is learned (e.g., facts, skills, etc.) and the quality of the learning (e.g., fluency)

Hughes uses the term 'processes' to cover any actions taken by the participants which might contribute to the process of learning, such as the development of materials, syllabus design, and teaching methods. Finally, 'products' refer to what is learned (facts, skills, etc.) and the quality of the learning (fluency, etc.). The trichotomy into participants, process and product allows planners to construct a basic model of backwash. Hughes (1993) suggests that the nature of a test may first affect the perceptions and attitudes of the participants towards their teaching and learning tasks.

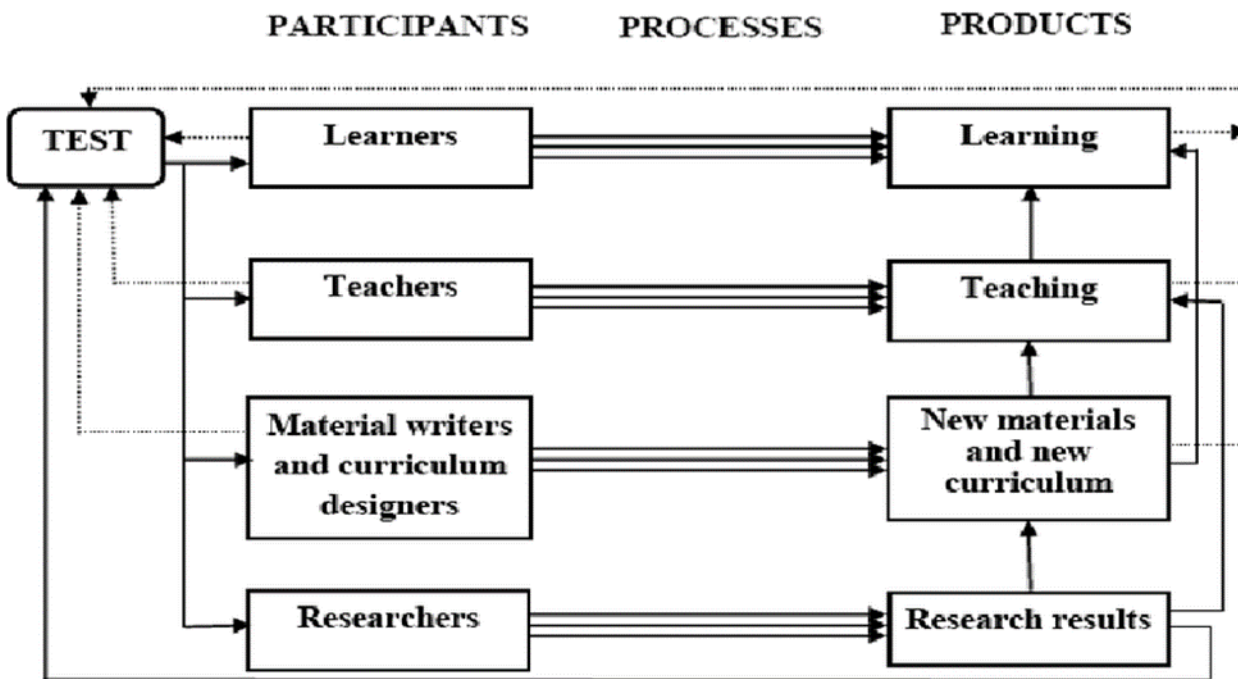
These perceptions and attitudes in turn may affect what the participants do in carrying out their work (process), including practicing the kind of items that are to be found in the test, which may affect the learning outcomes, the product of the work. As a pioneer model, it attempts to clarify how test works to desired outcomes. However, the model does not sufficiently clarify the term ‘processes’. As a first model of washback, it received worldwide recognition.

2.10.1.2 Bailey’s Washback Model

Based on Hughes’ (1993) tripartite distinction between participants ‘processes and products, Bailey (1996) develops and illustrates a model in which a test not only affects products through the participants and the processes they engage in, but where the participants and processes also in turn provide feedback and thereby also has an impact on the test, as dotted lines in Figure 2.1 represent. This model is an early attempt at theorizing washback but is not empirically grounded .

This model incorporates ideas from Hughes (1993) in describing a trichotomy of test effects in terms of “participants”, “process”, and “product”. Her model, however, is innovative in that it is grounded in empirical research evidence from educational change taking place in the Hong Kong context. Bailey points out participants include students, teachers, materials writers, curriculum designers, and researchers. Here, the participants refer to the stakeholders who directly participate in the teaching, learning, and testing process. Processes refer to the ways teaching is executed. Processes, according to Hughes (1993), refer to material development ‘syllabus design, changes in teaching methodology, and testing strategies among others. The products in a washback study refer mainly to what are learned and achieved. Products include learning, teaching, new materials and curricula, research results. Here, the focus is the development of communicative competence:

Figure (2.1) Bailey's washback model (1996)



Bailey's model is designed on the basis of suggestions of Hughes (1993); however, she does not clarify the process herself. Bailey's model (Figure 2.1) shows and describes the participants and products, but it does not give any information of process. An apparent shortcoming in this figure was that it showed a test directly influencing the participants, without articulating the role of beliefs held by the participants. In other words, the model did not explain why the participants did what they did. In addition, the model proposed by Bailey (1996) no longer finds strong support among researchers as a model of washback because it includes wider test effects such as those on teaching materials which can be referred to as impact, rather than being restricted to the effects that a test has only on teacher and learner behavior (i.e., washback) as defined by Hamp-Lyons (1997) and Wall (1997). However, her model has immensely contributed to the washback studies during the last decade. Her model can be considered as a gateway and one of the pioneer washback models for future researchers.

She listed ways that students who are about to take an important test may participate in any of the following processes (ibid.: 264-265):

1. Practicing items similar in format to those of the test.
2. Studying vocabulary and grammar rules.
3. Participating in interactive language practice (e.g., target language conversations).
4. Reading widely in the target language.
5. Listening to non-interactive language (radio, television, etc.).
6. Applying test-taking strategies.
7. Enrolling in test-preparation courses.
8. Requesting guidance in their studying and feedback on their performance.
9. Enrolling in, requesting or demanding additional (unscheduled) test-preparation classes or tutorials (in addition to or in lieu of other language classes).
10. Skipping language classes to study for the test.

Bailey contends that it is the selection from among these processes which could lead to beneficial or negative washback, "...depending on whether or not their use promoted the learners' actual language development (as opposed to their perceived progress or their test-taking skills alone) (1996: 265).

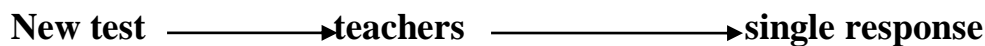
2.10.1.3 Burrow's Washback Models

Another set of simpler models is presented by Burrows (1998). As part of her doctoral study, she sought empirical evidence of the washback effect on the attitudes and practices of teachers on the Adult Migrant English Program in New South Wales in Australia. Her study looked at the impact of the implementation of the Certificate in Spoken and Written English. Her conclusions were that there was evidence of washback, but that different teachers reacted to the changes in assessment differently. She also felt that in her case, where testing and the curriculum were closely interwoven, the changes were not easy to separate.

Burrows (ibid.) identified three models of washback: one traditional pre-dating Alderson and Wall (1993); a second model, relating to current writing about washback e.g. Shohamy et al (1996); and she proposed a third model relating washback to curriculum innovation and teachers' Beliefs, Assumptions and Knowledge (BAK) as shown in the following diagrams (Figure 2.2):

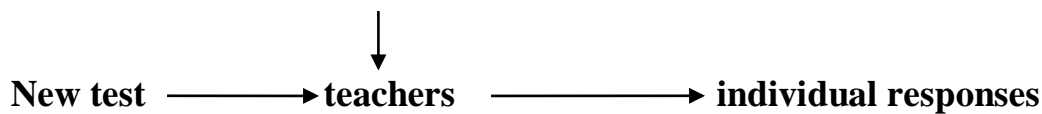
Figure (2.2) Burrow's washback models (1998)

Traditional Model



Current model

Bak



Model Proposed by Burrows

Bak



Burrows (ibid.) has argued that the models imply that a uniform and consistent washback effect would always be expected by the introduction of any test because the washback depends on the quality of the test rather than on the participants. She suggests that this early model is not based on objective evidence such as observation, but on teachers' anecdotal evidence. However, Burrows' models lack of discussion on the role of participants and teaching methodology. The models fail to draw wide attention of researchers due to their limitations.

2.11 Washback- a Phenomenon Leading to Educational Innovation

It is assumed that to understand the nature of washback, it is also crucial to take account of findings in the research literature in the area of innovation in language and change in educational settings. The reason of this view is that many applied linguists consider that there are many ideas in educational innovation which can stand as a solid ground for language testing specialists to judge whether the tests they are designing are likely to have the impact they intend them to possess Wall (2005). On this particular point, Hsu (2009) asserts that there has been a well-established tradition, which led to the realization of a number of networks that served to yield the most elegant compilations of assumptions about the different phases in the innovation process at the factors at work in every phases, and an increasing body of literature focusing on the English language teaching context. The particularity of these studies is that they managed to clarify the complexity of the innovation process, and the factors which inhibit or facilitate successful diffusion and implementation.

Following Wall (2005), defines innovation as an 'idea, practice, or object that is perceived as new by an individual or other unit of adoption'(p.11). In Hsu (2009), innovation can be usefully defined as a planned and deliberate effort, seen as a new act by an individual or individuals to bring about improvement in relation to desired objectives. Hsu makes this last assumption more explicit. He advocates that educational innovation is the result of a number of problems that a given educational system can present as a failure of students' achievement, a poor performance by students in specific Washback in Language Testing: A Review of the Concept and its Implications to Innovation in Education areas, or lack of transparent accountability reporting. What is significant about these problems is that the latter also transgress to touch some aspects of educational system that concern systematic attempts by some authorities to change educational policies with the intention to achieve better outcomes.

On the ground of this elucidation, a number of models have been provided to make the subject of innovation in language education more practical and possibly easy to incorporate in acts that intend to yield desirable changes or as it is proposed in the domain of testing, a factor leading for washback.

- The initiation stage: it is the process that occurs between the first appearance of the idea for change and the time when it is adopted.
- The implementation stage: it is the process of putting into practice an idea, program, or a set of activities and structure new to the people attempting or expected to change.
- The continuation stage: it refers to whether an innovation becomes part of the educational system, or whether it fails and/or is rejected.

In a nutshell, the purpose of eliciting this literature is to provide enough information on the intricate relationship between innovation in education and washback in language testing. Thus, it is evident that different theories of innovation and change have yielded insights on how researchers should provide to implement subjects that are new for the people concerned by this change. Besides, an understanding of the basics of these described models gives a better interpretation of the nature of washback, and more importantly, how this latter works when time comes to implement an innovative act.

2.12 Areas Affected by Washback

The studies discuss the effects of washback on various aspects of the classroom, which can be categorized as follows: curriculum, materials, teaching methods, feelings and attitudes, learning. The paper will review the findings for each of these areas in turn.

The view of testing is derived from the realization of test power and its manifestations with regard to high-stakes decisions based on test results for individuals, educational systems and society as a whole. Many research studies reveal that a test affects participants, processes, and products in teaching and learning.

Students, teachers, administrators, material developers and textbook writers may be included under the term ‘participants’. Their perceptions and attitudes towards their work are likely to be affected by a test. Process refers to any action taken by the participants, which may contribute to the process of learning .Material development, syllabus design, use of syllabus and curriculum, applying teaching methodology, and the use of learning and/or test-taking strategies are included under processes. Product means what is learned (facts, skills, etc.) and the quality of the learning (e.g. fluency, competence, etc.). Tests have an impact on the learning outcomes as well .As mentioned, washback affects various aspects of teaching and learning ‘such as syllabus and curriculum, stakeholders, materials, teaching methods, testing and mediating factors, learning outcomes, feelings, attitudes, and learning, etc. Tests have impact on the lives of test takers, classrooms, school systems and even whole societies Hamp-Lyons (1998). Wall & Alderson (1993) put forward the 15hypotheses, highlighting more specifically some of the ways in which a test might affect teaching and learning. The five of the hypotheses relate to washback to the learners, six relate to washback to the program, and four relate to syllabus ‘ curriculum, and teaching contents.

The Washback Hypothesis seems to assume that teachers and learners do things they would not necessarily otherwise do because of the test. Additionally, in order to study the washback effect, it is necessary to look at the people that participate in the educational process, to the actual classroom events and activities ‘and to the outcomes of these processes. Based on the various types of research throughout the world, washback hypotheses may be summarized as:

- 1 .Tests can affect curriculum and learning ‘
- 2 .Tests can provide feedback on learning ‘
- 3 .Tests can help implement content and performance standards ‘
- 4 .Tests can influence the methodology that teachers use ‘

- 5 .Tests can motivate teachers and students ‘
- 6 .Tests can orient students as to what is important to learn ‘
- 7 .Tests can help orient needed teacher training ‘
- 8 .Tests can help implement articulation ‘
- 9 .Tests can help implement educational reform .

A curriculum is a vital part of the EFL classes, and washback has deep relation with the syllabus and curriculum. Test contents can have a very direct washback effect upon teaching curricula. It provides a focus for the class and sets goals for the students throughout their study. A curriculum also gives the student a guide and idea to what they will learn, and how they have progressed when the course is over. The test leads to the narrowing of contents in the curriculum. Tests can affect curriculum and learning (Alderson & Wall, 1993). Shohamy et al. define curriculum alignment as “the curriculum is modified according to test results ”(1996, P.6). The findings from the studies about washback onto the curriculum indicate that it operates in different ways in different situations, and that in some situations it may not operate at all .

Learners follow a ‘hidden’ syllabus, that is, the contents driven by the contents of examination. Alderson and Wall (1993) conclude from their Sri Lank study that ‘the examination has had a demonstrable effect on the content of language lessons’ (p, 126-127). This effect is that of the narrowing of the curriculum to those areas most likely to be tested. This finding is similar to that of Lam (1994) who has reported an emphasis in teaching on those parts of the exam carrying the most marks. The findings of Read and Hayes (2004) are quite detailed and show variations in washback on the curriculum depending on the course observed. The studies discuss the effects of washback on various aspects of the classroom, which can be categorized as follows: curriculum, materials, teaching methods, feelings and attitudes, learning. This section reviews the findings for each of these areas in turn .

2.12.1 Washback on Syllabuses and Curriculums

Many researchers (e.g. Bailey, 1996, 1999; Wall & Alderson, 1993, Wang ‘2010 ‘ Hsu, 2009) of high-stakes tests attest that tests are responsible for narrowing the school curriculum by directing teachers to focus only on those subjects and skills that are included in the examinations. As a consequence, such tests are said to “dominate and distort the whole curriculum” (Vernon, 1956: 166; see also Kirkland ‘1971 ‘ Shepard, 1991). A test was considered to have beneficial washback, when preparation for it did not dominate teaching and learning activities narrowing the curriculum. When a test reflected the aims and the syllabus of the course, it was likely to have beneficial washback, but when the test was at variance with the aims and the syllabus, it was likely to have harmful washback .

Wall & Alderson (1993) put forward the 15 hypotheses, highlighting more specifically some of the ways in which a test might affect teaching and learning. The following are the hypotheses that relate to syllabus, curriculum, and teaching contents :

- (1) A test will influence what teachers teach; and
- (2) A test will influence what learners learn; and
- (3)A test will influence the rate and sequence of teaching (P); and
- (4)A test will influence attitudes to the content, method, etc. of teaching and learning (ibid) .

Examination should reflect the syllabus and curriculum, and since not everything in a curriculum can be tested in an examination, the areas that are assessed should be ones that are considered important. It is also important that, same items and contents should not be tested again and again. Insofar as possible, modes of testing (e.g., written, practical, oral) should be diverse to reflect the goals of curricula. The format and contents of the achievement test should be reorganized every year. The use of

commercially produced clone tests materials in the class should be discouraged. Teaching to the test universally occurs in either the practice of frontloading or backloading. If a high match exists between the curriculum and the test, teaching to the test is inevitable and desired. Otherwise, the data produced by the test is not useful in improving teaching and learning. In this case, using tests as the source to develop curriculum runs the risk of accepting and defining learning only in terms of what is tested in the test .

2.12.2 Washback and Teaching to the Test

Teaching to the test--the very words has always been heresy to educators . 'Teaching to the test' puts too much emphasis on standardized tests that are poorly constructed and largely irrelevant, the theory goes; it stifles creativity and encourages cheating. Vallette (1994) suggests that washback is particularly strong in situations where the students' performance on a test determines future career options. In such case, teachers often feel obliged to teach to the test, especially if their effectiveness as a teacher is evaluated by how well their students perform .

The assumption that frontloading alignment prevents teaching to the test is often not the case, in terms that teaching to the test still occurs under the practice of frontloading. If the curriculum and the test correspond to each other, teaching to the test is inevitable and desired. The extent to which a test is useful to a given curriculum is the extent to which the test indeed measures the curriculum in the first place. In the alignment by frontloading, examining the test itself is one way to assess the test quality, in terms of determining whether anything on the instrument that ought not to be taught is tested or that ought to be taught is not tested. A backloaded curriculum assumes "null curriculum"; that is, the content not tested or assessed in the test is not included in the curriculum. The act of "null curriculum" or "non-selection"

is valued laden. The values not selected by the test makers represent an unknown element that may be at odds with local values .

2.12.3 Washback on Teaching Methodology

By teaching methods, the present researcher refers to teaching approaches or techniques. The findings on this area are once again not homogeneous. While Alderson and Wall (1993, p. 127) says that their Sri Lanka study showed the exam ‘had virtually no impact on the way that teachers teach’. Andrews et al. (2002) point out that the revised exam led to teachers’ use of explanation of techniques for engaging in certain exam tasks.

Cheng (1997) mentions that teaching methods may remain unchanged even though activities change as a result of the revision of an exam; in this case reading aloud was replaced by role plays but both were taught through drilling (p, 52). The high-stakes EFL examination leads teachers to teach through simulating the examination tasks or through carrying out other activities that directly aim at developing exam skills or strategies (e.g., brainstorming, working in pairs or ingroups, jigsaw activities, simulating authentic situations, engaging in debates ‘discussions, speeches, etc.). Watanabe’s findings for this area are once again different. He reports that the 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’ (2000, p. 45) .

Some of the studies indicate that the methods used to teach towards exams vary from teacher to teacher. Alderson and Hamp Lyons (1996), find large differences in the way teachers teach towards the same exam or exam skill, with some adopting much more overt ‘teaching to the test’, ‘textbook slave’ approaches, while others adopted more creative and independent approaches (p, 292). The researchers in both these studies stress that the variable may be not so much the exam or exam skill as the teacher him=herself. They go on to discuss various teacher-related factors that may

affect why and how a teacher works towards an exam. Teacher attitude towards an exam would seem to play an important role in determining the choice of methods used to teach exam classes. There has been apperception that washback affects teaching content and teaching methods. It seems to be true in some circumstances but not others, suggesting that whether the exam affects methods or not may also depend on factors other than the exam itself, such as the individual teacher. Other findings on teaching methods relate to interaction in the classroom .

Alderson and Hamp Lyons (1996) note in their investigation of TOEFL teaching that the exam classes spend much less time on pair work, that teachers talk more and students less, that there is less turn taking, and the turns are somewhat longer. Watanabe (2004) notes that ‘students rarely asked questions even during exam preparation lessons’. Cheng (1998) points out that while teachers talk less to the whole class as a result of the revised exam, the teacher talking to the whole class remains the dominant mode of interaction .

It is seen that examination-oriented materials are heavily used in classrooms particularly when the examination approaches. However, it is not clear from the studies that it is the exam that generates less interaction in exam classes, or whether this is due to teachers believing, for whatever reason, that this is the way exams should be prepared for. The type and amount of washback on teaching methods appears to vary from context to context and teacher to teacher. It varies from no reported washback to considerable washback. The variable in these differences appears to be not so much the examination itself as the teacher .

2.12.4 Washback on Teacher Factors

Teacher perception, teacher attitudes and teacher beliefs are often mentioned in the washback studies as powerful factors. Among the factors that can mediate the washback effect is the teacher Wall (1996) and her/his perceptions about the

examination, its nature, purposes, relevance in the context, etc. What have been noted in the results are the behaviors of teachers in response to examination changes .

However, there is a need to examine the link between teacher intentions or beliefs and how this translates into action. Woods (1996). By doing so ‘predictable variations in teachers’ behavior that result from differences in goals ‘judgments and decisions can be better accounted for. According to Shulman (1986 ‘1987) research that links teachers’ intentions to their behavior provide a sound basis for educating teachers and implementing educational innovations .

It is argued that the dictates of achievement tests reduce the professional knowledge and status of teachers and exercise a great deal of pressure on them to improve test scores which eventually makes teachers experience negative feelings of shame, embarrassment, guilt, anxiety and anger. Green (2006, 2007) starts to examine this facet of washback. Johnson (1992), Sato and Kleinsasser (1999), Tan (2008)Turner (2006, 2008) and Wang (2008) have showed that teacher factors influence teaching practices in the classroom. Teacher beliefs are consistent with their prior experience and instructional approaches. There is, therefore, an increasing realization in the field of assessment that the “teacher factor” is fundamental to the kind of washback effect that takes place in the classroom .

Wall and Alderson (1993) comment the examination has considerable impact on the content of English lessons and on the way, teachers designed their classroom tests (some of this was positive and some negative), but it has little to no impact on the methodology they used in the classroom or on the way they marked their pupils 'test performance. Among many important results of the Sri Lankan impact study ‘Wall and Alderson make the following summary statements about the impact of the new Sri Lankan texts and tests on the teachers (ibid., p. 67) :

1 .A considerable number of teachers do not understand the philosophy/approach of the textbook. Many have not received adequate training and do not find that the Teacher's Guides on their own give enough guidance .

2. Many teachers are unable, or feel unable, to implement the recommended methodology. They either lack the skills or feel factors in their teaching situation prevent them from teaching the way they understood they should .

3 .Many teachers are not aware of the nature of the exam- what is really being tested. They may never have received the official exam support documents or attended training sessions that would explain the skills students need to succeed at various exam tasks .

4 .All teachers seem willing to go along with the demands of the exam (if only they knew what they were) .

5 .Many teachers are unable, or feel unable, to prepare their students for everything that might appear on the exam .

Watanabe (2004a) finds that the presence of grammar translation questions on a particular university entrance exam did not influence these two teachers in the same way. He has identified three possible factors that might promote or inhibit washback to the teachers: (1) the teachers' educational background and/or experiences; (2) differences in teachers' beliefs about effective teaching methods ‘and (3) the timing of the researcher's observations. (Teacher A was observed when the exams for which the students were preparing were six months away, while

Teacher B was teaching exam-preparation classes just a month or so before the entrance examinations would occur.) Thus, Watanabe concludes that "teacher factors may outweigh the influence of an examination" (ibid., p. 331) in terms of how exam preparation courses are actually taught .

Tests can aid learning and teaching both if aimed to assess the required skills. Many researches have been carried out on washback explicating that it can be either beneficial or harmful depending upon the contents and techniques (Alderson&Wall 1993; Bailey 1996, p. 257; Cheng &Falvey 2000). For example, if skills not required for every day communication are assessed, the test could leave harmful effect on

teaching and learning, such as mechanical test of writing skills by giving multiple-choice questions on grammar. A great number of washback studies focus on what takes place in the language classroom. Many researchers (e.g. Cheng 2004, Wall & Alderson, 1993; Turner 2007; Qi, 2004, 2005) find that content changes because of the test, but the way teachers instruct does not vary to any great degree. The changes were “superficial” Cheng (2005, p. 235), not substantial .

A majority of teachers tended to “teach to the test.” For example, Green (2007, 2006) and Hayes & Read (2003, 2004) find more test-related activities (e.g. Offering test-taking tips, doing question analysis) in the IELTS preparation classes than in the EAP (English for academic purpose) classes. In addition, teachers’ beliefs and attitudes regarding the immediate goals of teaching and their own limited ability to use the language effectively contribute to their being unable to effect the positive changes (a shift in English language teaching to a more communicative orientation) the test developers intended to create Qi (2005). Cheng (2004) asserts that inadequate training and teachers’ professional backgrounds lead to unchanged methodologies because they don’t know how to change, not that they do not want to change .

A good number of researchers find that tests affected both how and what teachers taught but not all teachers reacted the same way to the same test. In many instances, teachers reported a greater sense of pressure from the tests (Watanabe, 2004b; Burrows, 1998; 2004). Shohamy (1993); and Shohamy et al (1996) .also have discovered significant differences between experienced and novice teachers. The former tends to teach to the test and uses only material to be included in the test, while the latter uses different activities to teach oral language. Lam (1994) has reported that more experienced teachers tend to be significantly more “examination-oriented” (p. 91) than their younger colleagues. The new teachers are found more sincere language teachers than the experienced or older ones. The more the teachers

get experienced, the more the teachers teach to the test. The experienced teachers are relatively misguided by the examination, and thus, create very negative washback on their teaching .

The findings of the previous studies on teaching show that washback is contradictory in terms of what (content) and how (methodology) teachers teach. This may be attributed to Hawkey's claim (2006) that "the distinction between course content and methodology is not always clear cut" (p. 106). Nevertheless, researchers (Burrows, 2004; Cheng, 1997; Wall & Alderson, 1993; Watanabe, 1996; Watanabe (2004b) seem to have reached a consensus on the concept that tests influence what happens in the classroom in terms of teaching activities and content, and that teachers' beliefs, and educational backgrounds play an important role in deciding how they instruct the students in the class .

2.12.5 Washback on Language Learning

There is a general understanding that washback is a complex phenomenon .Many researchers call for empirical studies to explore the concept further. It is encouraging to note that more and more researchers have expanded to look at issues of context in order to capture the complexity of the washback phenomenon. It is obvious that the washback phenomenon has been examined much more seriously, both theoretically and empirically. In comparison to washback studies in other areas, fewer researches have been conducted to investigate the washback effects on students' learning processes. Watanabe (2004) states, "relatively well explored is the area of washback to the program, while less emphasis has been given to learners" (p. 22). Those studies that have been focused on learning washback received varied and sometimes contradictory findings .

2.12.6 Washback on Test Takers

The learners are the key participants whose lives are most directly influenced by language testing washback. The washback influences the test takers directly by affecting language learning (or non-learning), while the influences on other stakeholders will affect efforts to promote language learning. The test-takers themselves can be affected by: the experience of taking and, in some cases, of preparing for the test; the feedback they receive about their performance on the test ; and; the decisions that may be made about them on the basis of the test. Of the 15 washback hypotheses of Alderson and Wall's (1993, pp. 120-121), five are directly addressed learner washback. Bailey (1996) suggests students face with an important test they may participate in (but are not limited to) the following processes :

- 1 .Practicing items similar in format to those on the test .
- 2 .Studying vocabulary and grammar rules .
- 3 .Participating in interactive language practice (e.g., target language Conversations).
- 4 .Reading widely in the target language .
- 5 .Listening to non-interactive language (radio, television, practice tapes, etc.) .
- 6 .Applying test-taking strategies .
- 7 .Enrolling in test-preparation courses .
- 8 .Requesting guidance in their studying and feedback on their performance.
- 9 .Requesting or demanding unscheduled tutorials or test-preparation classes in addition to or in lieu of other language classes.
- 10 .Skipping language classes to study for the test. (pp. 264-265)

Learner washback has also important financial implications for pupils and their families, in terms of their access to educational opportunities. For example ,Wall and Alderson examined a context in which a new national test was implemented, this time the O-level exams administered at the end of the 11th year of education in Sri Lanka.

These authors report, "a student's O-level grades 'particularly in English, are among the most important in his or her academic career "(1993, p. 42). Washback may affect learners' actions and/or their perceptions, and such perceptions may have wide ranging consequences. Sturman used a combination of qualitative and quantitative data to investigate students' reactions to registration and placement procedures at two English-language schools in Japan. The placement procedures included a written test and an interview. He found that the students 'perceptions of the accuracy of the placement .

2.12.7 Washback on Materials

The term 'material' is used here to refer to the prescribed textbooks 'guidebooks and past question papers. Examination-related textbooks and other materials can vary in their type of contents. Very often, tests promote a boom of test related materials, and thus, influence what teachers teach in the classroom, but tests may also encourage teachers to use additional materials from a variety of sources .They range on the one hand from materials that are highly exam technique oriented 'and make heavy use of parallel exam forms, to those on the other hand that attempt to develop relevant language skills and language. A teacher's choice of materials relies on a number of factors such as the purpose of the test and the availability of ready-made materials. Generally, the studies refer particularly to those materials at the 'highly exam oriented' end of the spectrum .

A large number of studies discuss washback on materials in terms of materials production, the use of materials, student and teachers' views of exam materials, and the content of materials. Most teachers know from their own experience of the rows of exam-related materials available on the shelves of bookshops and staff rooms, and of the new editions of course books and other exam materials that are issued when exams are revised. They find that in relation to the EFL exam 'ample new material

has been published and marketed since the announcement of the test changes became public .

Teachers' use of materials seems to vary to large extent. Lam (1994) speaks of teachers as 'textbook slaves' and 'exam slaves' (p.91). He finds that large numbers of teachers rely heavily on the textbook in exam classes, and more heavily on past papers. Lam (1994) also reports that teachers do this, as they believe that the best way to prepare students for exams is by doing past papers. Andrews, et al (2002)) speak of the large role played by published materials in the Hong Kong classroom, citing a previous study by Andrews (1995) in which the teacher respondents were found to spend an estimated two-thirds of class time working on exam-related published materials. Cheng (1997) suggests that a reason for this maybe that the exam textbooks in Hong Kong not only provide information and activities but also suggested methods for teaching and suggested time allocations .

Andrews, et al. (2002) also speak of the large role played by published materials in the Hong Kong classroom, citing a previous study by Andrews (1995) in which the teacher respondents were found to spend an estimated two-thirds of class time working on exam-related published materials.

Tests that emphasize a communicative approach, such as the HSC often elicit heavy reliance on test-related materials by teachers. Teachers devote more attention to assisting students to achieve high scores rather than learn real communication skills. It may be, then, that in the viewpoint of teachers, using test-related materials can assist them in doing their jobs better in terms of helping students receive better scores. Tests promote a boom of test related materials and thus influence what teachers teach in the classroom, but tests may also encourage teachers to use additional materials from a variety of sources. A teacher's choice of materials relies

on a number of factors such as the purpose of the test and the availability of ready-made materials .

2.12.8 Washback on Lesson Contents

Learners follow a ‘hidden’ syllabus, that is, the contents driven by the contents of EFL examination. Many teachers, however, consistently skip over the listening lessons in their textbooks, because they know that listening will not be tested in the examination. A group of teachers may 'do listening', but in a way, that does not resemble the textbook designers' intentions. A few teachers cover the listening lessons if the type of question that students have to answer resembles an item type that might appear in the examination for reading. Most teachers in Bangladesh – particularly the higher secondary English school teachers, also admit they are influenced by the power of the public examinations. Thus, the status of their course is established by the importance of the teaching contents reflected on the entrance examinations .

There seems to be something of a mismatch between the attitudes of the teachers towards the contents of the learning package, and those of the students. The teachers clearly see the potential of the materials as a teaching package, containing relevant and worthwhile teaching activities, including but extending beyond test preparation. The students, on the other hand, are above all concerned with familiarizing themselves with the format of the test and seemed to be relatively little concerned with the learning strategies proposed, and the broader suggestions for improving performance .

In general, students demonstrate relatively little interest in the idea of using test preparation as an opportunity for language learning. Alderson and Wall (1993) conclude from their Sri Lanka study that ‘the examination has had a demonstrable effect on the content of language lessons’ (p, 126-27). This effect was that of the narrowing of the curriculum to those areas most likely to be tested. This finding is

similar to that of Lam (1994) who reports an emphasis in teaching on those parts of the exam carrying the most marks .

2.12.9 Washback on Learning Outcomes

Teaching to the test and test taking strategies might increase students' scores 'but the score gains are not always statistically significant. Moreover, class instruction of exam-specific strategies and non-class instruction factors such as students' initial proficiency, personality, motivation, confidence, and exposure of environment all possibly contribute to a score gain. A test itself does not lead to various aspects of the perceived effects. It is rather mediating factors such as teachers' beliefs and educational backgrounds, students' individual differences (e.g. motivation, English proficiency), and purpose of test use that play essential roles in causing test effects

It has been demonstrated that a test can result in all desired changes in teaching and learning. Wesche (1983), points out that when tests reflect the situations, content and purpose where learners will use the language, they are likely to improve motivation. Education is a complex phenomenon and there are many factors involved in bringing about changes, like the school environment, messages from administration, expectations of teachers and students, for example.

Wesdorp (1982) finds there is no difference in students' writing in quality before and after the introduction of multiple-choice tests. Hughes (1988) reports that at a Turkish university, students' performance on the Michigan Test (a measure of English proficiency) increases after the introduction of a new test along with additional summer courses in English. Andrews et al. (2002) investigate the score comparisons that students receive on the UE (Use of English) oral exam in Hong Kong from 1993 to 1995. Students' scores have increased, but the score gain is not statistically significant. They claim that students' improved proficiency might have something to do with their "familiarization with the exam format, the rote-learning of exam, specific strategies and formulaic phrases" (p. 220) .

In this sense, students' original proficiency plays a more important role in the resulting score gain than the time they spend in the test preparatory course. Score gain washback, as concluded from the foregoing discussion, is a complicated issue. It is difficult to detect what causes or does not cause it. Further research needs to be conducted to determine whether students have made progress because the test motivates them to study harder or if other factors such as their original proficiency, personality, motivation, and exposure have more weight in explaining the outcome .

2.12.10 Washback Impact on Society and Education System

The societal and educational value systems that inform the test use must always be considered by test users and test developers. The values and goals become very complex in the context of second or foreign language testing, since the values and goals that inform test use varies from different cultures. For example, one culture may emphasize individual effort and achievement, while another culture might emphasize group cooperation and respect for authority. Another aspect that needs to be considered is the consequences of our actions. We must realize that when we use a language test, it is likely to have consequences not just for the individual, but also for the educational system and society. This is of great significance when it comes to high-stakes testing (Bachman, Palmer, 1996).

In addition to this, McNamara (2000), discusses that the power of tests influences the reputation of teachers and schools, which could lead to a strong influence on the curriculum. McNamara states that ethical language testing should work to ensure positive washback from tests. However, sometimes the responsible authorities use assessment reform to drive curricular reform, believing that assessment can be designed to have positive washback on the curriculum.

2.12.11 Washback Impact on Feelings and Attitudes

Generally speaking, the studies note a gamut of rather negative attitudes and feelings generated by exams. Cheng mentions that students show mixed feelings towards the exam itself, recognizing on the one hand that the exam made them work to achieve good scores but at the same time thinking that exams were not an accurate reflection of all aspects of their study Cheng (1998: 296). She also speaks of the pressure felt by teachers, that teachers are worried about how the shy or less outspoken students will fare in the new exam, and of one teacher who admits she would feel guilty if she did not familiarize her students with the test formats. Once again, it seems that factors beyond the exam itself come into play in determining the amount and kind of washback. In this case they include teachers' attitudes and the stakes of the exam.

Studies of test anxiety and its facilitating or debilitating effects on both teachers and learners during the teaching and learning process merit further research as part of studies of washback. That exams impact on feelings and attitudes seems clear but how these in turn impact on teaching and learning is much less clear.

2.12.12 Washback on Resources

The studies mention that resources can affect washback. Factors mentioned are whether or not customized materials and exam support materials, such as exam specifications, are available to teachers and the types of textbooks available (Cheng, 1997; Hamp Lyons, 1998).

2.12.13 Washback on the School

Factors mentioned in relation to the school are as follows: its atmosphere and cultural factors such as learning traditions (Watanabe, 2000); how much administrators put pressure on teachers to achieve results (Smith, 1991; Shohamy et al., 1996); and the amount of time and number of students allocated to exam classes (Alderson and Hamp Lyons, 1996; Read and Hayes, 2003).

2.12.14 Washback on the Exam

The studies mention that various factors related to the exam itself can influence degrees and kinds of washback. These include: its proximity, its stakes, the status of the language it tests, its purpose, the formats it employs Shohamy et al (1996), the weighting of individual papers Lam (1994), when the exam was introduced and how familiar it is to teachers Andrews et al (2002). The factors focus on the individual teacher and on the teacher as part of a wider system. Teachers, like everyone else, operate in ideological, historical, economic and political contexts that affect their attitudes, beliefs and behaviors.

The studies to date do not show in what directions the factors push washback. For example, would a well-trained and educated teacher working with an exam of which he or she approved and about which he or she was well informed be more or less likely to adhere to the content of the exam in their lessons? The studies indicate that the answer to this question would likely be: it depends. There is also an interaction between the factors and between the factors and the teaching and learning contexts, which is not as yet described. The variety of the factors, their varying strength and the complexity of the interactions between them indicate strongly that washback does not always occur and that when it does it may do so in a variety of forms and intensities in different contexts.

2.13 Testing the Different Language Areas

In simple terms, a test is defined as a “method of measuring a person’s ability, knowledge, or performance in a given domain” Brown (2004, p.3). In other words, a test can be a set of techniques, procedures, or items that requires performance on the part of the test-taker. Tests must be explicit and structured, are usually relatively time constrained, and normally occur at identifiable times in a curriculum. Therefore, testing is an important part of every teaching and learning experience. It has to be prepared by the teacher who feels a need to improve their skills in constructing and administering classroom tests in English as a Second Language (ESL). Testing may

be primarily constructed as tools to strengthen learning and to motivate students as a means assessing the students' performance in the language.

Language is scientifically and linguistically defined as a system of systems. So, there are many areas relevant to the study of these different systems, such as: phonology, morphology, syntax, lexicography or lexicology, semantics, stylistics and pragmatics. In teaching language, the above-mentioned areas are considered as the major ones in addition to the four language skills (listening, speaking, reading and writing).

2.13.1 Types of Tests: -

Fundamentally to any discussion of testing is the purpose of the test. If we don't know why we are testing, then we probably shouldn't be tested in the first place. The appropriateness of a test in any given context will depend to a large extent on the reason the test is being given and the uses to which the test results may be put.

1. Major Types:

a. Achievement Tests: (Attainment Tests)

They refer to more formal tests that have been designed to show mastery of a particular syllabus or corpus (collection of written texts) of language. They are tests of developed skill or knowledge. They are standardized to measure skills and knowledge learned in a given grade level, usually through planned classroom instruction. They include end-of-term and end-of-year tests, end-of-course tests, school leaving examinations for certification, and other tests administered by external bodies (such as **University of Cambridge Local Examinations Syndicate UCLES**, **Key English Test KET** or Preliminary **EnglishTest PET**). They are similar to progress tests in that they are generally based on a syllabus and measure what has been taught and learnt. However, they are usually standardized and are likely to be similar from one year to the next (and the same in format). They are rarely constructed by the classroom teacher for a particular group of learners, and they are usually assessed on a **pass** or **fail**, or on a **grade basis**. Therefore, achievement test

scores are often used in an educational system to determine what level of instruction for which a student is prepared. Gleason, H.A. (1961).

b. Aptitude / Competence Tests:

They are the type of tests that predict how successful a learner will be in (for example) learning a foreign language. The rationale is that some people are better than others at learning languages, and aptitude tests aim to predict the learner's possible performance in learning a language (or that particular language). The most common aptitude tests set tasks that are related to the linguistic skills possessed by a good language learner (such as the ability to perceive sound and spelling correspondences and the ability to identify patterns in language).

c. Diagnostic Tests:

They enable teachers to identify any areas of weakness or difficulty, so that they can then plan and implement an appropriate remedial teaching program. They may be used to assess the knowledge and skills of learners in particular aspects of language before the start of a course (and consequently may be used for placement as well as course design purposes).

d. Placement Tests:

They are a special type of proficiency tests. They give an indication of the language levels of learners so that they can be placed in an appropriate class for English.

e. Proficiency Tests:

They are different in that they are not usually based on a particular syllabus but are used to measure achievement in relation to a specific (future) task that the candidate may be required to perform at a subsequent point of time. For example, may set out to determine whether the candidate has sufficient English to follow a course of Chemical Engineering for which the medium of instruction is English, or to do a job that requires the use of English. These tests rarely take into account any syllabus that a student may have followed. They are concerned with future potential performance

rather than past achievement. They are frequently administered to learners with varied language learning backgrounds.

f. Progress / Stop and Check Tests:

They measure how well learners have mastered the material (or skills) taught in class and (or) how well they have improved in a given area. Consequently, the focus is on short-term objectives (such as Unit Two: The Past Tense or Unit Four: Expressions of Future Time). These tests are usually written by the teacher to answer questions such as:

- Have the students learnt the target language well?
- Have I taught it effectively?
- Can we go on to the next part of the course? Gleason, H.A. (1961).

Progress tests are often given to **motivate** and reinforce learners.

2. Other Sub-types of tests:

a. Objective Tests: It's grading is independent of the person marking that test. There is usually a key of answers that leaves no room for subjectivity in grading. A typical example is multiple choice or false –true tests.

b. Subjective Tests: The score depends on the marker. A typical example is the test on free writing.

c. Speed Tests: They aim at measuring the speed of performance. The test is made a little longer than the time given. e.g. two hundred items on grammar to be answered in one hour.

d. Public Tests: They are prepared by a central authority and given on a country-wide scale. They are usually announced and relatively long and given at the end of a school cycle.

e. School Tests: They are locally prepared and given at the school level by the class teacher. They contrast with public tests in terms of length, scale, purpose and examiner.

f. Standard Tests: They are standardized and carefully designed and have undergone long experimentation and research. Each score has a special interpretation.

g. Normal Tests: They contrast with standardized tests. The majority of tests are normal.

h. Written tests: Answers are to be given in a written form.

i. Oral Tests: Answers are to be given orally.

j. Announced Tests: The teacher assigns the test material and fixes a certain date in advance.

k. Drop Tests: They contrast with the announced ones. They are given without previous announcement. They are usually short and aim at keeping students on the alert.

l. Classroom Tests: Questions are given and answered in class.

m. Home Tests: Questions are given in class but answered at home.

n. Closed-book Tests: Textbooks are closed while students are taking the test.

o. Open-book Tests: Students are allowed to use their books while answering the questions of a test. Heaton, J.B. (1988).

Note: For example, a test may fit in eight types at the same time. It may be an objective, achievement, local, normal, written, announced, classroom, close-book test without any inner contradictions of classification. On the other hand, it is obvious that some types exclude one another. For instance, a test cannot be announced and drop at the same time.

2.13.2 Language Testing: -

Language Testing is a kind of a school or any educational institutions assessment which refers to the specific procedures that the teachers and examiners employ / device trying to measure the language acquiring abilities. Testing can also play a major role in teaching i.e. a good test is essential for good teaching and learning process, especially in normal school programs.

- Purpose of Language Testing: -

Tests can serve as a process for many purposes such as:

- Identification of problems.
- Course grading.
- Assessment of teaching.
- Achievement. Measure students' achievement.
- Evaluation & Self-evaluation. Learners' and teachers' performance.
- Experimentation. Used in educational experiments.
- Promotion. Used to determine which students deserve to be promoted to higher levels.
- Reinforcement.
- Parents' information. Parents need to know their children progress.
- Placement / Leveling & Grouping. Grouping classifying students homogeneously into levels.
- Preparation and Remedy. Prepare for the course start design remedial programs for weakness.
- Diagnosis. Diagnosing problem areas.
- Prognosis. Predict whether a certain student can potentially succeed in a certain study program.
- Urge / motive. Many students study mainly because of tests and only when a test is announced
- Entrance / Admission. Many educational institutions use tests as entrance examinations (scores determine either acceptance or rejection).

- Types of Language Testing: -

Language tests can be classified and administered in many types and formats such as:

a. Traditional / Subjective Tests.

b. Objective Tests. e.g. Multiple-choice items.

c. Discrete (Separate/ distinctive) / Skills Tests. e.g. Listening / writing / reading / oral or speaking tests.

d. Global (comprehensive) Tests. e.g.

1. Dictation test: Students studied a text in class and then had to learn it at home before being given a dictation test on it

2. Cloze test: (The term is taken from Gestalt psychology) It is based on a passage from which every n^{th} word has been deleted. It is a deletion test that consists of a portion of text with certain words removed (cloze text) where the participant is asked to replace the missing words. It requires the ability to understand the context and vocabulary in order to identify the correct words or type of words that belong in the deleted passages of a text.

3. Intrusion test: It is based on addition rather than deletion. It is localized (requires a definition of scope), time constrained (it doesn't last forever) and authorized.

e. Competence Tests. e.g. Performance and interaction tests. They assess the individual's ability to use reasoning skills to solve problems. They measure reasoning skills required for administrative support positions and for certain operational positions. They can also be used to assist managers in making decisions for appointment and for placement in training program, identifying training and development needs, and counseling for career transitions.

2.13.3 The Language Areas to be Tested: -

Tests should cover all the language areas and skills involved in learning according to the objectives stated and designed by a specific learning program. The areas to be involved in testing are:

1. Grammar: Students are tested to measure their abilities to produce and understand grammatical structures.

2. Vocabulary: A vocabulary test aims at measuring the students' abilities to understand, produce and use the words of the target language.

3. Writing: It measures the students' abilities in:

- handwriting- to be evaluated at the early stages of learning language.
- spelling –to develop abilities to spell the words learnt.
- punctuation – ability to supply the omitted marks in a given text.
- composition – to measure abilities to write one paragraph or more on a certain topic.
- précis / summary – ability to comprehend, summarize and write a text within a given limit.
- translation – ability to understand a text in the native source language and then express and write the ideas in the target language.

4. Listening: It aims at measuring the students' abilities in:

- aural comprehension – to comprehend a passage / text through listening.

5. Reading: It aims at measuring students' abilities in:

- visual comprehension – to comprehend a passage / text through reading.

6. Oral / Speaking: A speech test aims at measuring students' abilities in:

- uttering / pronunciation – recognize and produce the language sounds.
- connected speech – recognize and produce stress and intonation as well as speak the foreign language correctly and in a reasonable speed. Gleason, H.A. (1961).

2.13.4 Testing the Language Areas: -

Each language area or skill may be tested in various types of tests and in many ways according to the different features and characteristic of each area. There for each area or skill should be dealt with and treated in testing in isolation with regard to the possible test type and format.

a. Testing Pronunciation:

A pronunciation test may include one or more of the following forms:

1. Reading aloud: A student may be asked to read aloud words, sentences, or passages that contain already taught material. This test may also evaluate intonation and stresses.

2. Auditory discrimination: Here the teacher pronounces a list of words in pairs while students are listening. The students then are required to distinguish the

identically (I) and differently (D) pronounced pairs. Such a test measures student, ability to recognize the phonemes of EFL.

3. Visual discrimination: It is similar to the auditory test. However, in visual the pairs are only read by the examinee, where as in auditory the pairs are only heard by the examinee.

4. Underlined graphemes: Here a group of words is given with a letter or a cluster of letters underlined in each word. The underlined letters are usually graphically identical but they are not necessarily so. All the underlined letters in the group are pronounced in the same way except one letter or clusters, which the students are required to recognized

e.g. *child* / *chimney* / *chin* / chemistry – in the word (chemistry) the italicized cluster is pronounced / k /.

5. Phonemic transcription: In advanced levels, a student may be asked to transcribe words, sentences, or passages phonemically or even phonetically.

b. Testing Grammar:

Grammatical structures may be tested in different ways, either to test recognition or production as follows:

1. Form modification: Here a bracketed word is required to be put in its correct form e.g. He (come) to school tomorrow.

2. Filling in spaces / blanks: A word, normally structural one, is missing from a sentence and is to be supplied.

e.g. They just been here.

3. Synthesis / linking: Two simple sentences or more are to be synthesized / linked into one compound or complex sentence. (using conjunctions).

e.g. – Ali broke his leg. He was playing football. (while)

- He studied hard. He failed the test. (although / in spite of, etc.)

4. Error location: Here students are asked to underline any grammatical mistakes in a sentence and rewrite it in its correct form.

e.g. She didn't cooked food yesterday.

5. Completion: One clause of a complex sentence is given and the other is to be supplied.

e.g.- In spite of the difficult test,

- If I had money,

6. Transformation: A sentence in a certain form is to be changed into another specified form.

e.g. – Subject into object pronouns / Present into past / Direct into reported / Active into passive, etc.

7. Multiple choice: It is often used to test recognition. Two or more answers are given for each test item, only one of which is correct. the student is to select and encircle the correct answer.

e.g. This the man arrived last night.

a. which b. who b. whom c. whose

8. Word arrangement: A group of words is to be arranged in a way that makes a grammatically acceptable sentence.

e.g. the / over / he / wall / jumped.

c. Testing Vocabulary:

Vocabulary tests may aim at evaluating the students' ability to produce certain words or recognize them and they may take the following forms:

1. Multiple choice: Here several answers are given to fill in a blank, the student is to choose the only correct answer to achieve recognition and not production.

e.g. He went to watch the football

a. competition b. game c. match d. contest

2. Synonyms: A certain word is used in a context and the student is required to give a synonymous word. e.g. This is the **chief** idea. (main).

3. Antonyms: A certain word is used in a context and its opposite is to be given according to the meaning of that word in that specific context.

e.g. This is an easy test. (difficult)

4. Word formation / derivation: A certain word in a sentence is underlined or a stem is given in brackets. The students are asked to supply the derivative correct form of that stem or underlined word.

e.g. – They have seen the **collect** of photos. (collection).

- (careful). He answered the questions

5. Matching: A list of words, phrases, clauses, statements or enquires is given and the students are to find the matching synonyms, antonyms, definition or responses.

e.g. – develop improve

- wrong right

- thank you not at all / don't mention it

- Can I use yours? Of course, / never mind / I'm afraid you can't.

6. Filling: a) A list of words is given and a group of sentences or a text missing the words in the list is provided as well. The students are to choose the correct word to fill in the space.

b) A content word is missing in a certain sentence and is to be supplied productively.

e.g. Hydrogen is a gas, but water is a

d. Testing Listening:

In testing listening, making an immediate distinction between pure listening or hearing test and listening comprehension test is needed.

1. Listening / hearing test: Here many students have difficulty in recognizing certain English sounds or discriminating between them or between different phonemes.

- e.g. students listen to four words and mark the ones that are identical (e.g. Minimal pairs features).

a. bar b. star c. car d. car

- Students listen to a whole sentence and mark the missing words in their test frame.

e.g. They can see the sun

a. line b. shine c. fine d. nine

- Listening test can go beyond minimal pairs frame.

e.g. She went to Paris when she was

a. four b. forced to c. forty d. fourteen

- Word stress. Students listen to poly-syllabic words read by the teacher and mark the main stressed syllable on their test paper.

e.g. progress / photography / intersecting / familiar

- Sentence stress. Short sentences should be given and students have to mark the stress syllable / intonation.

e.g. O.K. I'll be studying this later, not now.

2. Listening comprehension test: Here students initially read a text or a passage. Then, they listen to the text or part of it and answer multiple choice items or **WHQs** based on it.

e. Testing Writing:

Writing tests are classified into:

1. Testing handwriting:

- Small / capital letters – teacher pronounce and students write down. Here the student ability to recognize phoneme-grapheme association is evaluated.

- Imitation / Copy – A model sentence is given so as to be imitated repeatedly.

2. Spelling test:

- Dictation of selected words, phrases, sentences or short assigned passage.

- Roots and affixes (prefixes / suffixes). A root is given with a prefix or suffix to be added. The students are to combine the root and its affixes considering any probable changes. e.g. in + regular / country + s, etc.

- Error detection. A group of sentences with one or more words misspelled. The student is to underline the misspelled and write it / them correctly.

e.g. I gave my freind a present .(friend) .

- Missing letter. A word is given with one missing letter. Students have to supply the missing letter in. e.g. w _ ite / ans _ er etc.

- Multiple choice. Each group of words consists of several words, one of which is correctly spelt. The students have to encircle or underline the correct one. e.g. persieve / perceive / percieve / purcieve .

3. Punctuation test: Adding punctuation marks where needed:

- Students have to added a specific punctuation mark (stop /period – commas–colon/semicolon–inverted commas–exclamatory/question marks)
- Students have to add all the missing marks in a given text or paragraph.

4. Composition test:

- Writing on a certain topic freely.
- Writing on a specific topic after the topic-sentence of every paragraph is given to control thinking.
- Writing on a certain guided topic with quantitative limits assigned such as number of words, paragraphs, lines, or sentences.
- Unscramble sentences or paragraphs into a one meaning unified composition or text.
- Summarization. Writing a summary of a given text stating the main ideas, points and paraphrasing.

5. Translation test: It may make one of the following forms:

- A passage in foreign language to be translated into the native language.
- A passage in the native language to be translated into foreign language.
- Instead of the passage, separate sentences may be given to be translated.

f. Testing Reading:

When testing reading five levels of comprehension can be formed or distinguished:

1. Literal comprehension. (objective or realistic)

2. Reorganizational comprehension. (rebuild)

3. Infernal comprehension. (deductive)

4. Evaluative comprehension.

5. Appreciative comprehension.

In order to test reading, students focus on the information of a given text containing vocabulary relevant or close to their standard and after reading, they have to answer:

- Post reading questions (WHQs / Yes, No, Qs)
- Multiple choice items.
- True / false statements.
- Prediction and opinion / point of view enquiries (agree / disagree).

e.g. What do you think of

- Judgment questions. e.g. To what extent
- filling in sentences related to a comprehension passage presented with some content or factual words missing (not of grammatical nature because testing here is to evaluate understanding not mastering grammar). Students have to fill in the spaces according to the information offered in the passage.
- Pairing or match tables of incidents (given in the form of from clauses phrases, or words) from a reading passage.
- Ordering. A list of statements is given to students who are asked to arrange those statements in the order of their occurrence according to chronological facts of the passage.

g. Testing of Oral Skills / Speaking:

Oral skills are considered the most difficult to test or administer as well not being reliable. So, Oral skills are often tested subjectively because they have to be given individually in such forms:

1. One-to-one situation: The teacher asks or tests a student individually.

- listen and repeat. The teacher says words or sentences then the student repeats correctly.

2. Stress and intonation: The teacher asks a student to repeat or stress words or emphasize converting a statement into a question or surprise without lexical or grammatical changes.

e.g. He's coming. He's coming! He's coming?

3. Reading aloud: Teacher asks students to read a passage, text, or dialogue aloud as a test of pronunciation and general fluency.

4. Using pictures: The teacher can show a picture and asks students to speak about, describe, or comment orally. The teacher can ask questions about the picture and demand the students to give oral answers.

5. Speech test:

-The teacher can ask a student to talk about a familiar topic for a limited duration of time. The student is then evaluated for fluency and correctness.

- A student may be asked questions, each of which requires a brief answer of one sentence or two. The questions may be given by the teacher or through a tape with timed pauses between each question and the one that follows. The stimulus may be a picture, which the student is to describe or comment on orally.

6. Dialogue or conversation test: Here two students may be asked to participate in a conversation or a dialogue (a role play activity).

2.13.5 Making a Test Format: -

When a test is to be made or formatted, the examiner is expected to note the following principles:

1. It is very important to determine the test objectives (the language areas or skills to be evaluated), i.e. production or recognition.
2. Each item in the test must stick to those set objectives. Otherwise, the item loses validity, an essential characteristic of good testing.
3. The time allotted / allowed for taking the test should be adequate for answering the questions.
4. It is recommended to place easy items at the beginning of the test and let more difficult ones come later.
5. The test should include items varying in their degrees of difficulty in order to spread students over a fairly wide range of scores and thus discriminate between slow, average, and bright / fast students.

6. The teacher should in advance decide the weight of the test in relation to the final grade and inform students of this proportion. He should determine how he is going to grade the test. (weighing each item within the test).
7. Let the maximal grade of each question appear on the question sheet. This will help students vary emphasis and plan time distribution according to the relative weight of each question.
8. The test should be clear in its questions and instructions on how to answer. This is essential for the test reliability.
9. The test should minimize the role of blind guessing in score attainment.
10. The test should represent the greatest part of the assigned material so as to be fair and eliminate dependence on mere luck.
11. Penalty for wrong answers to discourage blind guessing should be planned beforehand and known to students before they start answering the test.

2.13.6 Giving a Test: -

The teacher has to consider these points while giving a test:

1. It is better to give some instructions before question sheets are distributed such as how and where to answer.
2. The rest of the instructions are given immediately after distributing question sheet.
3. Students should be given a few minutes to ask about unclear points related to the test.
4. No questions are allowed after the first five minutes if instructions are stated and understood clearly.
5. Students should receive no help during taking the test because teaching takes place before and after the test but not during the test.
6. Teachers should insist on students' honesty while taking the test and prevent cheating because cheating will dispossess tests of their functions and make them almost meaningless.

In order to minimize cheating, teachers are advised to try the following suggestions:

1. It is better to space students.
2. It is better to stand in front of the class with minimal movement.
3. Some students need to be reminded strictly right at the beginning that cheating will be severely penalized.
4. Cheaters must be strictly dealt with and penalized.
5. The teacher can make and use two parallel forms by including identical test items in content but different in wording.
6. The same test may be made into several forms by putting its items in different arrangement. Heaton, J.B. (1988).

2.13.7 The Characteristics of a Good Test: -

A good test is characterized by the following qualities:

1. Validity: A valid test actually measures what it claims to measure. If a test aims to evaluate spelling, it becomes invalid if it evaluates pronunciation. Validity may be classified into three types:

- a. Face validity. It implies the general look at the test format.
- b. Curricular validity. It is content validity. the should represent the subject matter (course material).
- c. Concurrent validity. It is statistical one, where students' scores on a certain test are compared to their scores on another test. Both tests scores are then processed into certain statistical formulas to obtain validity coefficient.

2. Reliability: A reliable test is one that is dependable. If the same test or a parallel one is taken again by the same students, the score average should be almost constant during a reasonable time length. e.g. if the average of students' scores on a certain test is 80% and the average of the re-test scores drops down to 40% during a week's time then the test is not reliable.

Reliability may be classified into:

- a. Scoring reliability. It means that marking the same answers is consistent when they are re-marked even by other teachers.
- b. Temporal / test-re-test reliability. The same test is taken again by the same students after a certain period of time long enough to let students forget how they responded to the items in the first test but short enough to minimize probable additional learning.
- c. Internal / split-half reliability. splitting the test items into two halves – items of odd numbers and items of even numbers. A score for each half is given. The correlation between the two sets of scores is calculated and the coefficient is judged to be significant or not. There should be a high positive correlation between the scores on the two halves.

The teacher can do the following to increase reliability:

- Giving clear instructions on how to answer the different test items.
- Eliminating the luck factor in obtaining scores (controlling the nature of questions and increasing the number of choices in recognition items).
- The test has to be reasonably long.
- Cheating is to be absolutely forbidden and strictly dealt with.
- Eliminating blind guessing by giving a penalty for wrong answers and a correction of statements recognized to be wrong should be required.

3.Scorability: A scorable test should allow easy and accurate marking without wasting too much time or effort.

4. Representativeness: The test has to include items from different areas of the materials assigned for.

5. Discrimination: The test should distinguish among the different levels of students. e.g. if the test scores are between 90% and 80% then it has failed to show the students' individual differences (the test is very easy) and if the scores are between 20% and 30% it indicates that it is very difficult.

6. Time: Inadequate (much less time) or over-adequate (much more time) durations usually result in very low or very high scores respectively.

Gleason, H.A. (1961).

2.14 Part Two: Previous Studies

Introduction

This module presents previous studies related to the current study that have been conducted in different parts of the world (international, regional and local)

1.Sahar Ameer Bakhsh, (2016). *“Testing Communicative Language Skills of the Speaking Test in EFL Classrooms at King Abdul-Aziz University”*

This essay will first briefly identify other methods of testing before communicative language testing emergence, including what they measure and their theoretical basis. Next, an examination of the models on which communicative language testing is based on and ways in which communicative testing differs from other forms of language testing will be put forth. Then, an example of the speaking test administered in the preparatory year EFL classes at KAU to determine whether they meet the criteria they claim to test is going to be discussed. Finally, some of the problems communicative language testing faces and how these problems have been addressed will be examined.

2. Muhammad Asim Mahmood, (2014). M. A *“Impact of Public Examination on Teaching of English: A Washback Perspective Department of Applied Linguistics, GC University, Faisalabad.”*

This paper investigates the washback effect of Higher Secondary School Certificate (HSSC) exam on English language teaching and learning in Pakistan. The purpose of this study is to explore the causes of low level of English language proficiency at the completion of 12 years of academic education with English as a compulsory subject. To achieve this goal, it examines the effect of HSSC exam on teaching methods and materials. Population of the study was all the teachers teaching English in colleges at HSSC level in Sargodha, Pakistan. A survey questionnaire consisting of ten questions

was administered to the teachers of English. The questionnaire aimed at investigating the effect of HSSC on English teaching methodology and materials used by the teachers. Findings of the study affirmed that HSSC instead of supporting has negative washback effect on teaching methods and materials. Keywords: washback, HSSC exam, teaching method, teaching materials

3. Anne Dragemark Oscarsson, (2014). *“What makes a good language test in EFL?”*

This paper reports on what research says regarding the various components that are needed when constructing and using a language test. The findings point towards the importance of validity, reliability, and washback and the fact that these issues should be addressed with high consideration in order for a test to have a positive effect. We can see that evidence points to the fact that when tests are used, they have to measure what they are supposed to measure and that the evidence in validity is crucial. Furthermore, the terms test-retest and parallel tests were emphasized when discussing the reliability concept even though those methods have problems. Moreover, when the concept of washback was examined, it was clear that it is a powerful tool for both language learners and teachers. The literature suggested that the focus should be on impact and not processes. Finally, the presented criticism towards certain language tests showed that the tests were not used to assess language proficiency and had both reliability and validity issues. As it seems, most classroom tests are neither very reliable nor possibly valid because teachers are not able to construct proper tests with all these features. The results of this review seem to indicate that there is a lack of research regarding on how this gap could be closed and therefore deserves more attention

4. Sima Sadeghi, (2014). *High-stake Test Preparation Courses: Washback in Accountability Contexts*

The study was conducted with four instructors teaching at the preparation courses for TOEFL and IELTS in Iran. An interpretive ethnographic case study was conducted through observation and field notes to gather data about how high-stakes testing

affected teachers' curriculum and methodology. In order to keep observation focused and directed, University of Cambridge Observation Scheme (UCOS) was also used. The scheme was utilized to solely focus on the extent to which the course was test-oriented. The results indicated that teachers were constantly encountered with challenging questions which led to variations in their responses to tests' pressures .

5. Anthony Green, (2014). Ph.D. *“The Test of English for Academic Purposes (TEAP) Impact Study: Report 1 - Preliminary Questionnaires to Japanese High School Students and Teachers. Centre for Research in English Language Learning and Assessment (CRELLA), University of Bedfordshire, UK”*

This report relates to the introduction of the Test of English for Academic Purposes (TEAP), an innovative four-skills test of English for use as a university entrance exam in Japan. A key objective of the TEAP is to encourage good practice in the teaching and learning of English in line with the national course of study, giving due weight to spoken as well as written language. However, research suggests that good test design alone is not sufficient to bring about intended changes in educational practices. Action must also be taken to inform and educate stakeholders .

Advancing positive impact has been a part of the TEAP design and development process from the first. A program of research has been planned to investigate the impact of the TEAP and how far the test helps to bring about the intended changes in practice .

6.Chanika Gampper. (2013). *“Improving English Test Qualities. Language Institute Thammasat University”*

English Tests are widely used in classrooms and outside. Many of them draw a lot of criticism for not being an accurate measurement tool while their results are used to make important decisions for stake holders.

This article looks into the definition of a test, its five necessary qualities, namely reliability, validity, authenticity, backwash, and practicality and how to improve them so that English tests are good, effect, and useful.

7. Anthony Green, (2013) *“Washback in language assessment. University of Bedfordshire”*

This paper reviews the progress made in washback studies over the quarter century since Hughes’ (1989) placed it at the center of his textbook *Testing for Language Teachers*. Research into washback and the development of models of washback are described and an agenda is suggested for test developers wishing to build washback into their programs. It is recommended that future projects should pay greater attention to test design features and to the outcomes of learning as well as continuing to explore learner motivation and cultural factors that might encourage participants to react to tests in certain ways, but not in others. Washback research itself is seen to be a potentially valuable tool in persuading participants to adopt new practices.

8. Robert Kirk Patrick, (2012). Ph.D. *“Washback Effects of the New English Assessment System on Secondary Schools in Bhutan, Gulf University of Science and Technology, Kuwait”*

This study examines the washback effect of the English assessment system associated with this new curriculum. Based on questionnaire responses from 56 EFL secondary school teachers in Bhutan, the study suggests the new curriculum produces both positive and also negative washback. In order for the new curriculum’s benefit to outweigh any harm it may cause, teachers should receive proper guidance and support during this transition.

This paper discusses the effects of the new education system, its curriculum, and the assessment systems and includes: (1) the history of Bhutanese education system and its policy goals, (2) a summary of the new English curriculum for secondary schools in Bhutan, (3) the washback effects of the new English assessment system on Bhutan secondary schools, and (4) recommendations on how to make the assessment more effective

9. Md. Enamul Hoque, (2011). *“Washback of the Public Examination on Teaching and Learning English as a Foreign Language (EFL) at the Higher Secondary Level*

in Bangladesh. Department of English Jahangirnagar University Savar, Dhaka, Bangladesh”

The key objectives of the study were to examine washback as a phenomenon relating to those areas that were most likely to be directly affected by the HSC examination in English. The study set out a number of research questions and answered them to achieve the objectives of the study. The whole study is presented in this thesis divided into six chapters, each chapter incorporating specific issues of the present study.

10. Esma Şenel - Birsen Tütüniş, (2011)

“THE WASHBACK EFFECT OF TESTING ON STUDENTS’ LEARNING IN EFL WRITING CLASSES”

This study displays the literature review that investigates what the drawbacks of exams are on learning in general, then gives a brief account of a small-scale research conducted in two EFL writing classes. There has been extensive research in related literature on the impact of testing on students learning a foreign language. In practice most of the teachers assess their students with scheduled exams which they conduct in accordance with the syllabuses prepared at the beginning of the semester. The level and form of the negative effects change according to some factors such as gender, age, the aim and level of learning. This study investigated the negative effects of testing and their reasons on elementary level EFL students’ writing. The results indicate that learners’ anxiety level increase and their performance decrease when they know that it is an assessment.

11. Marian Amengual Pizarro, (2010). *“Exploring the Washback Effects of a High-Stakes English Test on the Teaching of English in Spanish Upper Secondary Schools, University of Illes Balears”*

This paper focuses on the washback effects of a high-stakes English Test (ET) on the teaching of English. The main goal of this study is to examine the washback effects of the ET on the following aspects of teaching: curriculum, materials, teaching

methods, and teaching feelings and attitudes. The study also attempts to discover teachers' perceptions towards the introduction of a speaking and a listening component in the design of the new ET due to be implemented in 2012. The overall findings, collected from a questionnaire carried out among 51 secondary teachers, indicate that the ET is clearly affecting curriculum and materials. Results also reveal that the ET appears to influence teachers' methodology. Furthermore, most of the teachers believe that the introduction of a speaking and a listening component in the new ET design will help solve the mismatch between the communicative approach they seem to value and the skills so far evaluated in the ET.

12. Yi-Ching Pan, (2009). Ph.D. “A review of Washback and its Pedagogical Implications” The university of Melbourne, Australia

This study starts with a focus on the various definitions of backwash or washback. Next, it examines the similar concept terms defined by other researchers. By reviewing the variety of definitions, the researcher's own view of washback will be reached. Comes after that are the explorations of different types of washback. The study ends with drawing pedagogical implications for EFL teachers.

13. Mohammad Reza Ghorbani, (2008). “The Washback Impact of the Iranian University Entrance Examination on Pre-University English Teachers”

This study examines the nature and scope of the washback impact of the university entrance examination (UEE) on pre-university English teachers' (PETs) perceptions. Teaching experience, educational background, professionalism, and gender were studied as independent variables. Based on stratified random sampling, 377 PETs were selected to respond to the questionnaires. Eight PETs were also purposively selected to participate in two focus group interviews. Pearson product moment and t-test were used to analyze the quantitative data from the survey questionnaires and a systematic note-based technique recommended by Kruger (2002) was used to analyze the qualitative data from the focus group interviews. The findings show that only PETs' perceived professionalism in teaching was associated with their perceptions.

All of the interviewed PETs perceived the negative effect of the UEE and expected the authorities to reform it based on the current teaching and testing theories.

14.Aliye Karabulut, (2007) M. A “*Micro level impacts of foreign language test. A thesis submitted to the graduate faculty in partial fulfillment of the requirements for the degree of*
MASTER OF ART”

The purpose of this study is to find out whether the foreign language examination - university entrance test- influences the way teachers teach and students learn in senior three classrooms (the last grade of high school) in Turkey. Secondary goal is to see the outcomes of teaching to the test and attitudes of different stakeholders towards the test and senior three English teaching in general .The data were collected through online surveys, and participants comprise of four major groups. Senior three high school students and English teachers were invited to participate to find out the nature and the scope of washback, while college students and professors are asked to participate to investigate the outcomes of teaching to the test .Descriptive statistics were used to analyze the responses of the participants. The results suggest that the test is a major factor determining the flow of English lessons in senior three classrooms. The classroom materials that were reported by both students and teachers including mock tests, commercial exam preparation materials and sample test questions directly serve to the purpose of practicing for the test and indicate the relative effect of the test on language learning .The results also suggest that high school students and teachers focus more on the immediate goal of language learning which is to score high on the test and be admitted to the university by cramming for the test and learning and practicing the language areas and skills that are measured on the test (grammar, reading, vocabulary) and ignore the ones that are not tested (listening, speaking, writing). Professors and college students, on the other hand, feel the shortage of not having enough practice especially in productive skills. They opine that long-term goal of language learning should be to improve the ability to use the

language .Based on the gap reported by these different stake-holders, findings lead to recommendations for a change in the curriculum and in the format of the test towards a more communicative and integrative one.

15.Mary Spratt. (2005). *“Washback and the classroom: the implications for teaching and learning of studies of washback from exams. ELT consultant, Cambridge, UK”*

This paper reviews the empirical studies of washback from external exams and tests that have been carried out in the field of English language teaching. It aims to do so from the point of view of the teacher so as to provide teachers with a clearer idea of the roles they can play and the decisions they can make concerning washback. The paper begins by defining its use of the term ‘washback’, then goes on to identify the areas in which washback has been noted by the studies. It next examines what intervening factors the studies have indicated influence whether and to what degree washback occurs. This examination highlights how much washback cannot be considered an automatic or direct effect of exams. Finally, the paper pulls together suggestions from the washback literature on how to teach towards exams and indicates areas of classroom practice that these could be applied to. The paper shows how crucial a role the teacher plays in determining types and intensity of washback, and how much teachers can therefore become agents for promoting positive washback.

16.Jahanbakhsh Nikoopou, (2005). *“Washback in Language Testing. A Comparison of the Entrance English Major Test of State and Islamic Azad University: A Wash Back Study”*

This study was based on the basic model of wash back proposed by Hughes (1993). The participants being directly influenced by the Newly Developed English Major Test (NDEMT); that is, high school teachers and those students who attempted the NDEMT were the subjects of his study. Two questionnaires and an observation checklist were used to investigate the wash back effect of the NDEMT on teachers

and students. The prime aim of this research put into the following ten hypotheses is to find out whether the NDEMT has wash back on Iranian EFL candidates and high school EFL teachers. (1) The subjects do not agree on the correspondence between the content of high school English courses and that of the NDEMT. (2) The subjects do not agree on the correspondence between the instructional practice of high school English courses and the types of English tests used in the NDEMT. (3) The NDEMT does not have any wash back on the content of the educational program, the evaluation methods of high school English instruction, the test construction and administration procedure, the types of test used in high school classes, the way test results are analyzed, the teachers' performance, the students' attitude toward it, and the teachers' attitude toward it. The findings of the study showed that all these null hypotheses were rejected. In addition, the results were almost the same in state and Islamic Azad University. Moreover, the positive and negative wash back effects of the NDEMT were put forward. Some strategies were suggested to promote positive wash back effects and some for reducing the harmful ones .

17. Belinda M. Hayes, (2003). *“IELTS PREPARATION IN NEW ZEALAND: AN INVESTIGATION INTO THE NATURE OF THE COURSES AND EVIDENCE OF WASHBACK.*

A thesis submitted to the Victoria University of Wellington in fulfilment of the requirements for the degree of Doctor of Philosophy in Applied Linguistics. Victoria University of Wellington”

This study investigated the washback effect of the test by studying three IELTS preparation courses offered by language schools at public tertiary institutions in Auckland. The aim was to identify the significant activities in an IELTS preparation class in New Zealand and establish whether there was evidence of washback in the way classes were designed and delivered. Various forms of data-gathering were utilized, including two structured observation instruments, questionnaires and interviews for the teachers, two questionnaires for the students, and pre- and post-

testing of the students. In addition, an analysis was made of IELTS preparation textbooks, with particular reference to those which were sources of materials for the three courses. Thus, the study provided a detailed account of the range and duration of activities occurring in IELTS preparation courses as well as insight into the teachers' approach to selecting appropriate lesson content and teaching methods.

The findings showed markedly different approaches between the courses, with two focusing almost exclusively on familiarizing students with the test and providing them with practice on test tasks. On the other hand, the third course, while including some test practice, took a topic-based approach and differed from the others in the amount of time spent on the types of activities one might expect to find in a communicative classroom. Pre- and post-testing revealed no significant gain in overall IELTS scores during the courses .

18.Liyng Cheng, (2000) Ph.D. “A review of the impact of testing on teaching and learning. Queen’s University”

This paper aims to share the discussion of this education phenomenon from different perspectives both in the area of general education and in language education. It discusses the historical origins of washback; the definition and scope of washback; and the function and mechanism of washback, and efforts, both recent and not, to mitigate its negative effects. It is concluded that the ultimate reason for the persistence and widespread nature of this problem is the existence of high-stakes testing. Few educators would dispute the claim that these sorts of high-stakes tests markedly influence the nature of instructional programs. Whether they are concerned about their own self-esteem or their students' well-being, teachers clearly want students to perform well on such tests. Accordingly, teachers tend to focus a significant portion of their instructional activities on the knowledge and skills assessed by such tests.

(KFT)

19.Glenn Fulcher, (2000). *“The ‘communicative’ legacy in language testing. English Language Institute, University of Surrey, Guildford, Surrey GU2 7XH, UK”*

This article looks at the phenomenon of ‘communicative’ language testing as it emerged in the late 1970s and early 1980s as a reaction against tests constructed of multiple choice items and the perceived over-emphasis of reliability. Lado in particular became a target for communicative testers. It is argued that many of the concerns of the communicative movement had already been addressed outside the United Kingdom, and that Lado was done an injustice. Nevertheless, the jargon of the communicative testing movement, however imprecise it may have been, has impacted upon the ways in which language testers approach problems today. The legacy of the communicative movement is traced from its first formulation, through present conundrums, to tomorrow's research questions.

20.Kathleen M. Bailey, (1999) *“Washback in Language Testing. Educational Testing Service Princeton, New Jersey”*

This monograph summarizes recent research on language testing washback. It begins by compiling several definitions of washback and related constructs. It then poses a model of language testing washback and examines the available research related to this model. The monograph concludes with recommendations for appropriate research methods to be used in future investigations of washback.

2.15 Summary

This chapter is concerned with the review of literature of the theoretical framework of washback, origin of washback, the definition of washback, its connection to impact, positive and negative connotations, the function and mechanism of washback. In addition to, the impact of washback on the participants, the areas and the factors that influenced by the process of washback, as well as the review of testing the different language areas. The pervious researches and studies in the field of this study have been stated by the researcher at the end of this chapter.

From this review of the literature, we can see that language testing washback (1) has often been discussed; (2) is widely held to exist; (3) that there are differing points of view about what the construct may encompass; and (4) that positive washback is viewed as an important criterion in the development and evaluation of language tests. However, until recently very little empirical research has investigated the phenomenon in detail.

CHAPTER THREE

METHODOLOGY

3.0. Introduction

This Chapter discusses the research design and methodology employed in this study. The study has adopted the descriptive analytical method. Two tools were used as data collecting methods (test for students and a questionnaire for teachers). It focuses on the method, population and sampling the procedures of data collection as well as statistics and the summary.

3.1 Methods of Analysis (Statistics)

The researcher utilized descriptive and statistical methods in which frequencies, percentages, arithmetic averages and their standard deviations are used. The results are sectioned and discussed with reference to the research questions. The Statistical Package for the Social Sciences (SPSS) was used to analyze the collected data and (APA) style was adopted to describe the basic features of the required data in a manageable form. After organizing and summarizing the collected data in a sensible way, Pearson product moment correlation was then applied to analyze them. The collected data was transformed into figures and tables to facilitate interpretation.

3.2. Population of the Study

The population of the present study consists of first secondary school students in the Eastern Region of Saudi Arabia. In addition to, the teachers of English as a foreign language in the same region.

3.3. Sample of the Study

The sample of the study consists of (30) students of first secondary school for the administration of the test. The students were homogenous in terms of age, learning experience and abilities. That is, most of the students are in range of (16- 18) years of old. They have been studying English language for six years, and (30) teachers of

English as a foreign language (EFL) were selected for the administration of the questionnaire to check the influence of washback-based approach to the achievement test on assessing the learning process.

3.4 Tools of Data Collection

The field study is mainly carried out by implementing two tools to collect and analyze the required data for this study. The first tool was a written (pretest and posttest) which was given to first year secondary students to measure the subject's ability towards the impact of the test results intelligibly and influentially on developing the EFL learning process. The second tool was a questionnaire which was given to teachers of English of English as a foreign language in the Eastern Region of Saudi Arabia to support the questions and the hypotheses of the research.

3.5 Test Design and Format

After a review of test design techniques and study of books of test design and format, taking in consideration the syllabus taught to the population of the study as well as the website sources, the researcher started developing the test format which consists of five sections to investigate the learners' competence in recognition of the impact of washback on EFL testing in the field of reading, grammar, vocabulary, conversation and writing. The test is then developed to include the most important characteristics of a good test in order to insure its validity and reliability.

Before putting the test in its final draft, the items were carefully reviewed and adjusted by handing them over to a group of skillful specialists so as to make relevant amendments (see appendix C). The researcher seeks their comments. In the light of their comments, the test was put in its final draft.

The test items are intended to measure the students' responses in concern with the impact of washback-based approach of the achievement test in developing learning process. The test covered the following domains (Table 3.1) of the EFL teaching, learning and testing conceptions.

Table (3. 1) Components and items of students' pre and post tests

No.	Components	No. of Items
1.	Reading	2 Items
2.	Structure and Grammar	2 Items
3.	Vocabulary	2 Items
4.	Conversation	1 Item
5.	Writing	1 Item

3.5. 1 Test Procedures:

The researcher adopted the following procedures to tackle the problem of the study:

- 1.Specification of the study population and sampling from Qatif Intermediate School students. The sample number is (30) students.
- 2.Collecting and investigating of students' marked test papers. One grade is given for each correct answer and a zero for an incorrect one.
- 3.Identify the current errors related to reading, grammar, vocabulary, conversation and writing
4. The results of the pre/post tests were registered, compared and analyzed with reference to the research questions

3.5.2 Test Administration

Before administering the test, the researcher briefly told the students of the study about the purpose and significant of the test, assured them that their answers would be treated confidentially and their test results would affect neither their academic achievements nor their obtained marks in the examination.

3.6 Teaching Plan (Treatment)

At the beginning of the academic year (2016- 2017), the researcher designed a specific teaching plan aiming at improving first year secondary students' language knowledge and enable them to use the language communicatively or properly in the different context of life and manage their learning process, since it was observed that students were showing weakness in this respect.

For carrying out the plan, all available techniques, materials, and aids were exploited. However, before students being exposed to this plan, the researcher exposed the target sample to a diagnostic test (pre-test) in order to check the actual problem and need of students' language knowledge. Soon after the pre-test administration and registration of the results, the researcher started carrying out the plan systematically.

After four months the same test was administered to the same group for the second time as an achievement test (post-test) and the results were registered. The results of the two tests (pre/posttests) were compared to see whether the teaching plan had a significant effect on the students' performance. Methodically the whole design of experiment is called one group pre/ posttest design or paired samples. For more experimental control, the factors such as age, culture, learning experience and others were taken into consideration to strengthen the effect of the treatment.

3.7 Test Statistical Treatment

To demonstrate the hypotheses of the researcher, the researcher counted the frequencies with their percentage and put them into tables to show which errors are most common among the study population and to figure out the most commonly used strategies.

3.7.1 Test Validity

In order to check the apparent validity for the study test and validation of its questions according to the formulation and explanation, the test was checked by four

Ph.D. holding referees in the study field (see appendix C). Some of the referees made suggestions, and others confirmed that the test was suitable. In any way, the researcher studied their suggestions, and some corrections have been made.

3.7.2 Test Reliability

The test is reliable when it gives consistent result if it is reapplied in the same conditions Brown and Rogers (2002: 241). The researcher establishes the test reliability by employing the Retest procedure which involves giving the test twice to the same group of students which concluded that there is higher correlation between the scores of each student in the two administrations of the test.

Statistical Reliability of the Test

Cronbach's Alpha	N of Items
.89	10

The above table shows the reliability of the test which was 0.89. That means if we redistribute the test gain the percentage of the same results.

3.8 Questionnaire Design and Format

The questionnaire was constructed by taking care to cover all the aspects relating to washback- based approach of the achievement test.

The final version of the questionnaire consisted of two main parts. The first part consisted of seven categories of teacher personal characteristics related to the personal information, including name, job title, place of work, age, gender, academic and professional qualifications.

The second part consisted of six categories and 45 items altogether. This part mainly dealt with current teaching and learning situations, the medium of instruction in the classroom and the teachers' attitudes and perceptions of the impact of washback on EFL testing. The statements were closed-ended items in different issues. All six categories of this part were designed on a three-point Likert-type scale of agreement

and the statements range from 'Agree', 'Neutral' to 'Disagree'. The survey was conducted during the academic year 2016/ 2017. The completion of the questionnaire took approximately two weeks. The teacher questionnaire dealt with EFL domains of teaching, learning and testing conceptions.

The teachers were directed to put a tick mark (√) in the box next to each statement, which best expresses their opinions.

Part 1: Background and personal information of the respondents:

In this part, the teachers were requested to give their answers thoroughly by ticking what suits them in the provided spaces.

Part 2: The questionnaire:

This part includes the sections and statements that directed to the teachers of EFL in order to give their answers to six up-to date conceptions in the field of teaching and learning process that directly related to the impact of washback on ELF achievement test system.

The teacher questionnaire (Appendix-B) was consisted of (45) items and structured in (6) sections.

Section 1 aimed at statements about the syllabus and curriculum such as curriculum objectives, teaching the syllabus, skipping lessons and feeling pressure to cover the syllabus, etc.

Section 2 consisted of a set of statements related to the textbook English for today and other materials used in the class.

Section 3 included statements on teaching methods and classroom behaviors.

Section 4 were on the classroom tasks and activities that usually took place in the class.

Section 5 included the statements on skills and linguistic elements of EFL usually practiced by them.

Section 6 consisted of statements on the attitudes, beliefs, and perception towards the achievement test . The questionnaire covered the following domains (Table 3.2) of the EFL teaching, learning and testing conceptions.

Table (3. 2) Components and items of teacher questionnaire

No.	Components	No. of Items
1.	Syllabus and Curriculum	7 Items
2.	Textbooks and Materials	10 Items
3.	Teaching Methods	9 Items
4.	Tasks and Classrooms Activities	5 Items
5.	Language Skills and Elements	5 Items
6.	Students' Attitudes and Perceptions Related to the Test and Teaching	7 Items

3.8.1 Questionnaire procedures

The researcher adopted the following procedures to tackle the problem of the study:

1. Specification of the study population and sampling (30) teachers of EFL in the Eastern Region of Saudi Arabia.
2. Identifying the current statements that ticked by the teachers and completely related to the current teaching and learning situations. Like, curriculum, syllabus, textbook, material, classroom task and activities, EFL skills, beliefs and attitudes.
3. Distributing a questionnaire survey to (30) EFL teachers who were requested to give their professional and technical inputs and points of views in concern with the statements that completely related to washback-based approach of the achievement test on the learning process.
4. Collecting, investigating and analyzing the frequencies and the percentages of the questionnaire.
5. The results were sectioned and discussed with reference to the research questions.

3.8.2 Questionnaire Administration

Before administering the questionnaire, the researcher briefly told the subjects of the study about the purpose and significance of the study, assured them that their responses would be treated confidentially and their answers would affect neither their academic achievements nor their obtained marks in the examinations.

3.9 Questionnaire Statistical Treatment

To demonstrate the hypotheses of the researcher, the researcher counted the frequencies with their percentage and put them into tables to illustrate the results and to show which statements are the most common among the study population and to figure out the most commonly used strategies.

3.9.1 Questionnaire Validity

By examining the validity for the study questionnaire and validation of its statements according to the layout and illustrations, the items of the questionnaire were judged by four Ph.D. holding referees who were specialists in the study field of English. (see appendix C). Necessary modifications and correction have been recommended by jury in both form and content of the questionnaire positively, and the final version of the instrument was adopted by the researcher. Thus, the questionnaire can be considered valid.

The value of the reliability and the validity lies in the range between (0-1). The validity of the questionnaire is that the tool should measure the exact aim, which it has been designed for.

In this study the validity calculated by using the following equation:

$$\text{Validity} = \sqrt{\text{Reliability}}$$

For calculating the validity and the reliability of the questionnaire from the above equation, the researcher distributed (30) questionnaires to respondents to calculate the reliability coefficient using the Alpha-Cronbach coefficient; the results have been showed in the following table

3. 9.2 Questionnaire Reliability

For checking the instrument reliability of questionnaire, Spearman's coefficient of correlation formula was used in order to find out the reliability coefficient

Here are some of the most used methods for calculating the reliability:

. Alpha-Cranach coefficient

The reliability coefficient was calculated for the measurement in the questionnaire was Alpha-Cronbach coefficient Equation.

For calculating the validity and the reliability of the questionnaire from the above equation, the researcher distributed (30) questionnaires to respondents to calculate the reliability coefficient using the Alpha-Cronbach coefficient; the results have been showed in the following table

Reliability Statistics

Cronbach's Alpha	N of Items
.83	47

3.10 Summary

This chapter has discussed the test and the questionnaire design format, test procedures, test administration, the teaching plan or the treatment, the statistical treatment, research methodology and the research tools adopted for data collection. The chapter has provided a detailed description of all the steps and procedures followed in each tool, including population, sample, validity and reliability of each tool.

CHAPTER FUOR

DATA ANALYSIS AND DISCUSSION

4.0 Introduction

This chapter presents the analysis of data collected through pretest, posttest and questionnaire then discusses the results to provide answers to the questions and hypotheses of the study. The researcher has utilized of the descriptive statistical method in the first step and statistical analysis of the test and the questionnaire to code and analyze data collected. It is worth mentioning that data analysis is conducted by using the SPSS program which means "statistical package for social science."

4.1 The Responses to the Students' Pre- Post Tests

Before the treatment, a pretest (*PET test*) was given to the target sample in order to guarantee their homogeneity and determine their ability and knowledge. The test which consisted of 5 questions separated in different parts and the total marks is 50 was administrated to the same group. The students' results in the pre-and post- tests are listed in following table.

Table (4.1): The Subjects' Results in the Pre-and Post-Tests

Student No.	Questions Items										Total 50 Marks	
	1		2		3		4		5			
	Reading 10		Grammar 10		Vocabulary 10		Conversation 10		Writing 10		Pre	Post
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post		
1-	4	8	6	6	5	7	5	10	0	3	20	33
2-	6	10	5	5	5	10	6	10	0	4	22	39
3-	5	9	5	8	4	5	5	10	4	5	23	37
4-	8	10	7	7	5	7	5	10	5	8	30	42
5-	7	9	3	5	4	7	3	9	0	3	17	33
6-	9	10	3	5	1	6	0	4	0	3	13	28
7-	5	10	7	7	8	10	8	10	5	7	33	44
8-	3	6	1	3	4	5	1	4	0	5	9	23
9-	4	6	1	3	0	2	0	3	0	2	5	16
10-	3	7	3	8	4	5	0	6	0	3	10	29
11-	5	10	5	10	10	10	5	10	5	7	30	47
12-	9	10	2	4	0	3	0	4	0	6	11	27
13-	8	10	8	10	8	8	10	10	5	5	39	43

Student No.	Questions Items										Total 50 Marks	
	1		2		3		4		5			
	Reading 10		Grammar 10		Vocabulary 10		Conversation 10		Writing 10		Pre	Post
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post		
14-	8	10	10	10	10	10	10	10	8	9	46	49
15-	4	7	2	5	0	3	0	5	0	3	6	23
16-	9	9	4	6	3	5	8	10	5	6	29	36
17-	8	9	0	3	5	5	10	10	0	2	23	29
18-	7	8	4	4	6	7	3	4	5	5	25	28
19-	8	10	2	5	9	9	10	10	5	7	34	41
20-	8	10	8	9	7	8	8	10	5	7	36	44
21-	6	7	6	6	5	9	5	10	3	4	25	36
22-	5	8	2	4	7	8	0	2	0	3	14	25
23-	4	4	1	3	1	5	0	3	0	2	6	17
24-	5	10	5	7	8	8	5	10	0	2	23	37
25-	4	6	0	4	0	3	0	0	0	2	4	15
26-	6	6	0	2	0	2	0	3	0	3	6	16
27-	7	10	3	5	3	7	5	5	2	4	20	31
28-	4	9	6	6	5	8	5	6	3	4	23	33
29-	8	10	8	9	10	10	9	10	8	9	43	48
30-	7	7	4	5	7	8	10	10	5	7	33	37

4.2 Analysis and Discussion of Students' Test

The students' pre and post/tests results were calculated and classified according to (mean, SD, DF, T-value- p-value) for hypothesis testing purposes

Table (4.2): T-test for the differences between the scores in pre and posttests in reading part

Tests	Means	SD	DF	T- value	p-value
Pre –test	6	0.21	29	14	0.040
Post –test	9	0.2			

For the scores gained from the pretest (*PET test*) the mean value calculated. was (6), and the mean value for the post test was (9) Moreover T-test was employed on these scores for hypothesis testing purposes. As the result of T-test suggests (P-value 0.040) being smaller than (0.05), our null hypothesis is rejected, this indicate that There is significance difference in students' scores after we retesting them on reading part.

Table (4.3): T-test for the differences between the scores in pre and posttests in grammar part.

test	Means	SD	DF	T- value	p-value
Pre-test	4	0.4	29	16	0.0401
Post test	6	0.3			

For the scores gained from the pretest (*PET test*) the mean value calculated. was (4), and the mean value for the post test was (6) Moreover T-test was employed on these scores for hypothesis testing purposes. As the result of T-test suggests (P-value 0.0401) being smaller than (0.05), our null hypothesis is rejected, this indicate that There is significance difference in students' scores after we retesting them on grammar part

Table (4.4): T-test for the differences between the scores in pre and posttests in vocabulary part.

Group	Means	SD	DF	T- value	p-value
Pre-test	5	0.33	29	14	0.02
Post test	8	0.43			

For the scores gained from the pretest (*PET test*) the mean value calculated. was (5), and the mean value for the post test was (8) Moreover T-test was employed on these scores for hypothesis testing purposes. As the result of T-test suggests (P-value 0.02) being smaller than (0.05), our null hypothesis is rejected, this indicate that There is significance difference in students' scores after we retesting them on vocabulary part

Table (4.5): T-test for the differences between the scores in pre and posttests in conversation part.

Test	Means	SD	DF	T- value	p-value
Pre-test	5	0.29	29	15	0.00
Post test	9	0.4			

For the scores gained from the pretest (*PET test*) the mean value calculated. was (5), and the mean value for the post test was (9) Moreover T-test was employed on these scores for hypothesis testing purposes. As the result of T-test suggests (P-value 0.00) being smaller than (0.05), our null hypothesis is rejected, this indicate that There is significance difference in students' scores after we retesting them in conversation section

Table (4.6): T-test for the differences between the scores in pre and posttests in writing part.

Test	Means	SD	DF	T- value	p-value
Pre-test	2	0.2	29	14	0.0212
Post test	4	0.3			

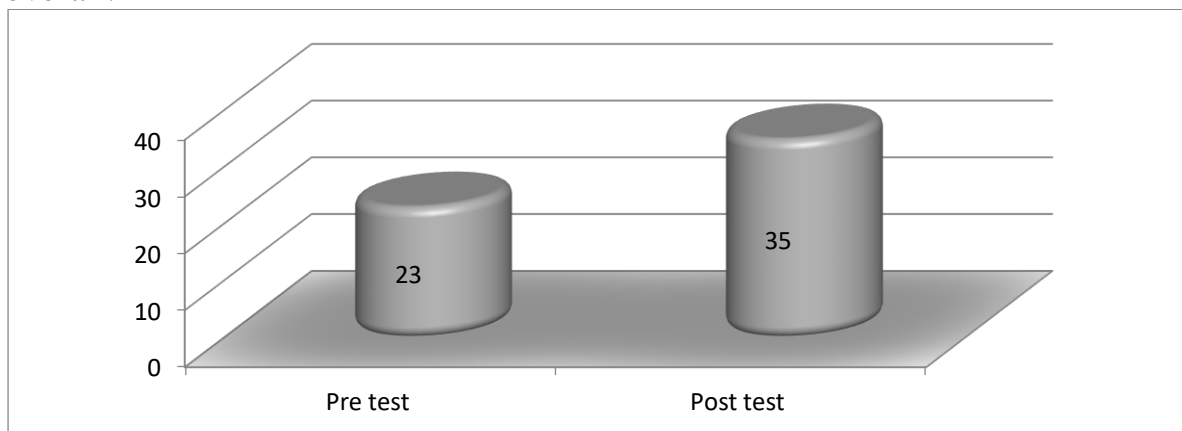
For the scores gained from the pretest (*PET test*) the mean value calculated. was (2), and the mean value for the post test was (4) Moreover T-test was employed on these scores for hypothesis testing purposes. As the result of T-test suggests (P-value 0.0212) being smaller than (0.05), our null hypothesis is rejected, this indicate that There is significance difference in students' scores after we retesting them on writing part

Table (4.7): T-test for the differences between the scores in pre and posttests for overall.

Test	Means	SD	DF	T- value	p-value
Pre-test	23	0.05	29	14	0.0000
Post test	35	0.33			

For the scores gained from the pretest (*PET test*) the mean value calculated. was (23), and the mean value for the post test was (35) Moreover T-test was employed on these scores for hypothesis testing purposes. As the result of T-test suggests (P-value 0.0000) being smaller than (0.05), our null hypothesis is rejected, this indicate that There is significance difference in students' scores on the overall part.

Figure (4.1) T-test for the differences between the scores in pre and posttests for overall.



It is obvious from the above figure the scores gained from the pretest (*PET test*) was (23), and the scores gained from the post test was (35). Therefore, our null hypothesis is rejected and this indicates there is significance difference in students' scores on the overall part.

4.3 The Responses to the Teachers' Questionnaire

The responses to the questionnaire of the (30) teachers were tabulated and computed. The following is an analytical interpretation and discussion of the findings regarding different points related to the objectives and hypotheses of the study.

Each item in the questionnaire is analyzed statistically and discussed. The following tables and figures will support the discussion.

4.4 Analysis and Discussion of Teachers' Questionnaire

The researcher distributed the questionnaire on determined study sample (30) and constructed the required tables for collected data. This step consists transformation of the qualitative (nominal) variables (agree, neutral, disagree,) to quantitative variables (1, 2, 3,) respectively, also the graphical representations were used for this purpose.

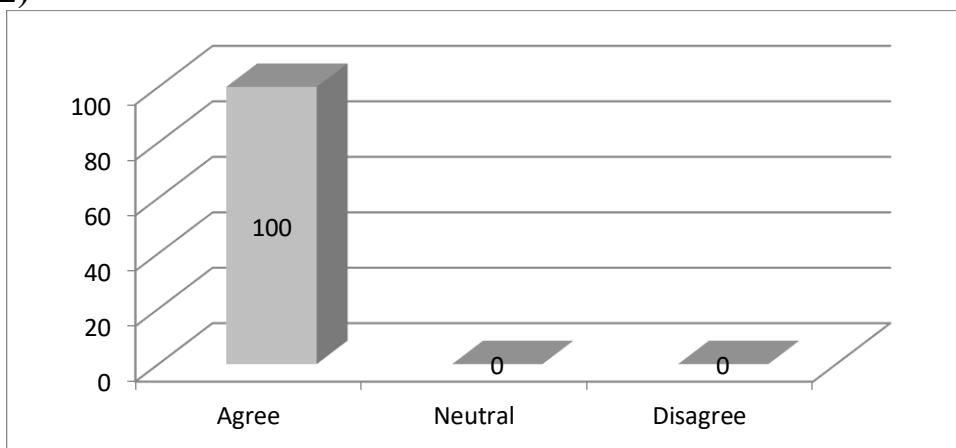
A. Curriculum and Syllabus

Statement No. (1): Teachers are aware of the objectives of the syllabus and curriculum.

Table No (4.8) The Frequency Distribution for the Respondent's Answers of Statement No (1)

Valid	Frequency	Percentage
Agree	30	100
Neutral	0	0
Disagree	0	0
Total	30	100%

Figure (4.2)



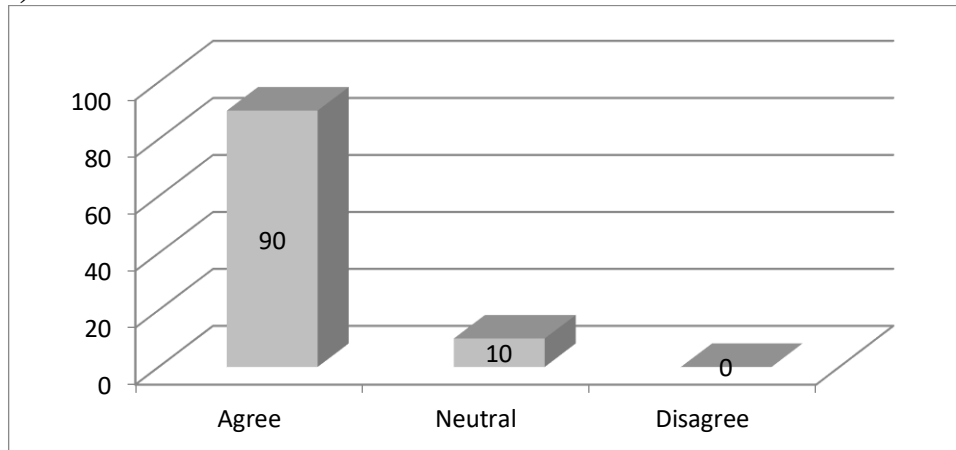
It is obvious from the above table and figure that there are (30) respondents in the study sample with percentage (100%) agreed with that "Teachers are aware of the objectives of the syllabus and curriculum. ". and (0) respondents with percentage (0%) were neutral and (0) respondents with percentage (0%) disagreed.

Statement No. (2): The present syllabus and curriculum can enhance EFL learning process

Table No (4.9) The Frequency Distribution for the Respondents' Answers of Statement No (2)

Valid	Frequency	Percentage
Agree	27	90
Neutral	3	10
Disagree	0	0
Total	30	100%

Figure (4.3)



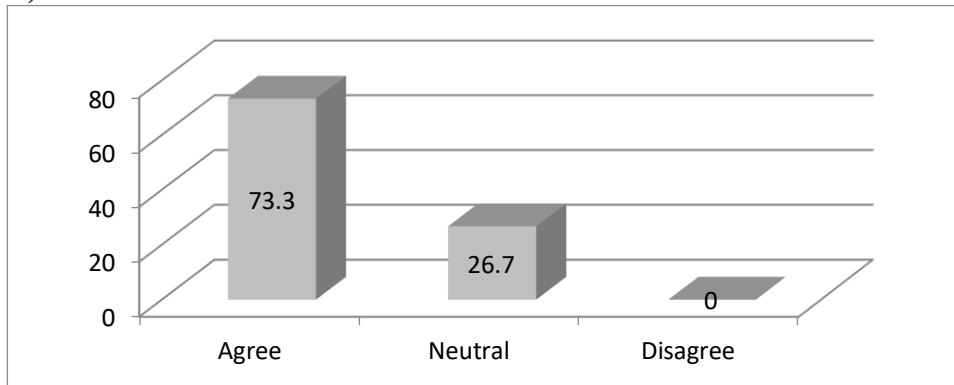
It is obvious from the above table and figure that there are (27) respondents in the study sample with percentage (90%) agreed with that "The present syllabus and curriculum can enhance EFL learning process". and (3) respondents with percentage (10%) were neutral and (0) respondents with percentage (0%) disagreed.

Statement No. (3): Teachers teach every section in the textbook (*Traveller*) although some sections are unlikely to be tested in the examination.

Table No (4.10) The Frequency Distribution for the Respondents' Answers of Statement No (3)

Valid	Frequency	Percentage
Agree	22	73.3
Neutral	8	26.7
Disagree	0	0
Total	30	100%

Figure (4.4)



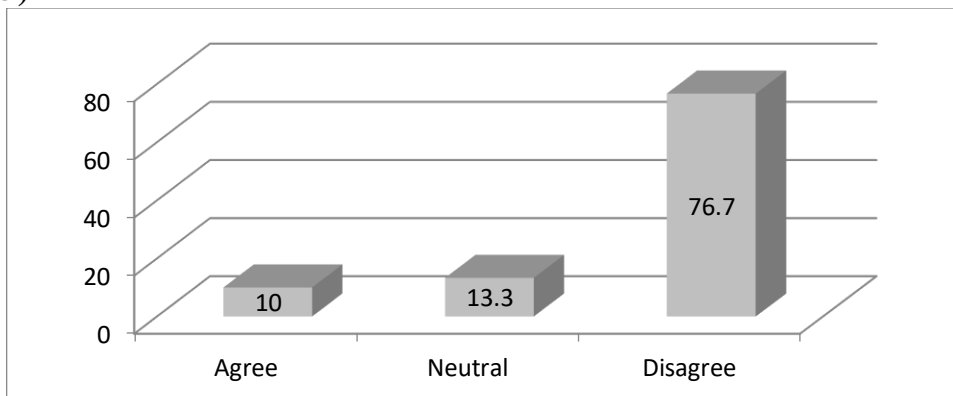
It is obvious from the above table and figure that there are (22) respondents in the study sample with percentage (73.3%) agreed with that “Teachers teach every section in the textbook (*Traveller*) although some sections are unlikely to be tested in the examination”. and (8) respondents with percentage (26.7%) were neutral and (0) respondents with percentage (0%) disagreed.

Statement No. (4): Teachers do not care about the syllabus and curriculum while teaching their students.

Table No (4.11) The Frequency Distribution for the Respondents’ Answers of Statement No (4)

Valid	Frequency	Percentage
Agree	3	10
Neutral	4	13.3
Disagree	23	76.7
Total	30	100%

Figure (4.5)



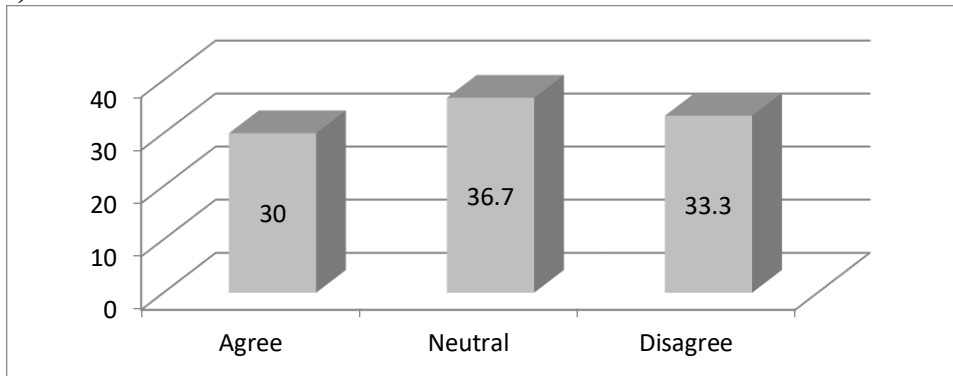
It is obvious from the above table and figure that there are (3) respondents in the study sample with percentage (10%) agreed with that "Teachers do not care about the syllabus and curriculum while teaching their students.". and (4) respondents with percentage (13.3%) were neutral and (23) respondents with percentage (76.7%) disagreed.

Statement No. (5): Teachers feel pressure to cover the syllabus before the final test.

Table No (4.12) The Frequency Distribution for the Respondents' Answers of Statement No (5)

Valid	Frequency	Percentage
Agree	9	30
Neutral	11	36.7
Disagree	10	33.3
Total	30	100%

Figure (4.6)



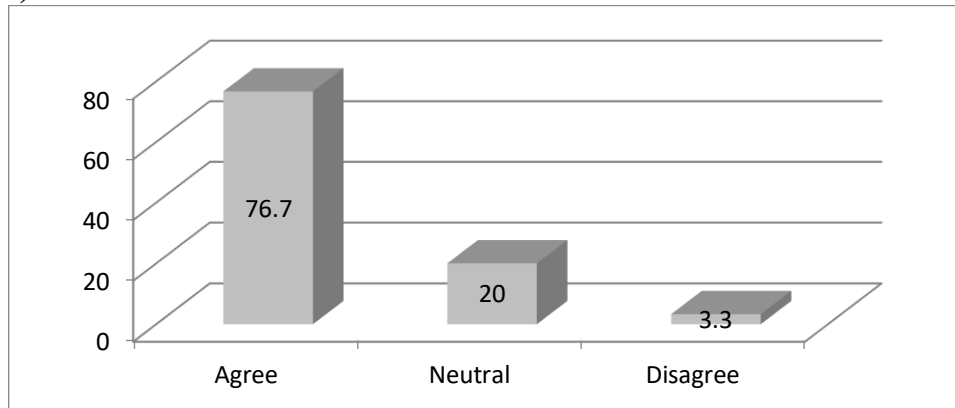
It is obvious from the above table and figure that there are (9) respondents in the study sample with percentage (30%) agreed with that "Teachers feel pressure to cover the syllabus before the final test.". and (11) respondents with percentage (36.7%) were neutral and (10) respondents with percentage (33.3%) disagreed.

Statement No. (6): The test tests the overall competence of students in English (listening, speaking, reading, and writing).

Table No (4.13) The Frequency Distribution for the Respondents' Answers of Statement No (6)

Valid	Frequency	Percentage
Agree	23	76.7
Neutral	6	20
Disagree	1	3.3
Total	30	100%

Figure (4.7)



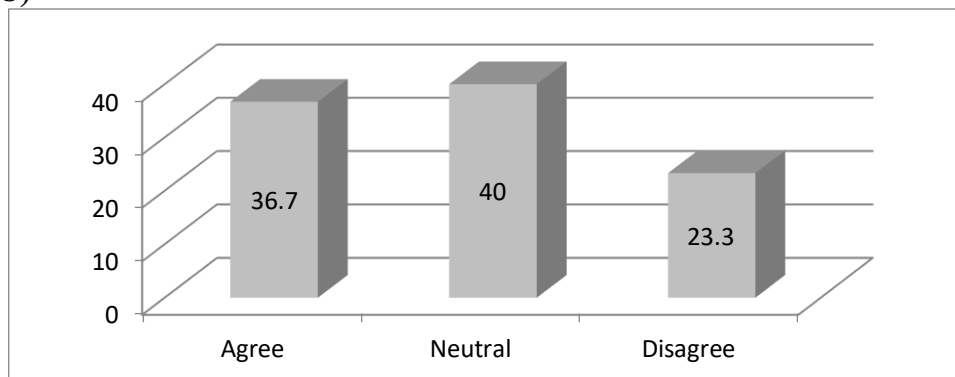
It is obvious from the above table and figure show that there are (23) respondents in the study sample with percentage (76.7%) agreed with that “The test tests the overall competence of students in English (listening, speaking, reading, and writing). ”, and (6) respondents with percentage (20%) were neutral and (1) respondents with percentage (3.3%) disagreed.

Statement No. (7): Teachers give more attention to teaching to the syllabus opposed to practicing the test items.

Table No (4.14) The Frequency Distribution for the Respondents’ Answers of Statement No (7)

Valid	Frequency	Percentage
Agree	11	36.7
Neutral	12	40
Disagree	7	23.3
Total	30	100%

Figure (4.8)



It is obvious from the above table and figure that there are (11) respondents in the study sample with percentage (36.7%) agreed with that “Teachers give more attention to teaching to the syllabus opposed to practicing the test items.). ”, and (12) respondents with percentage (40%) were neutral and (7) respondents with percentage (23.3%) disagreed.

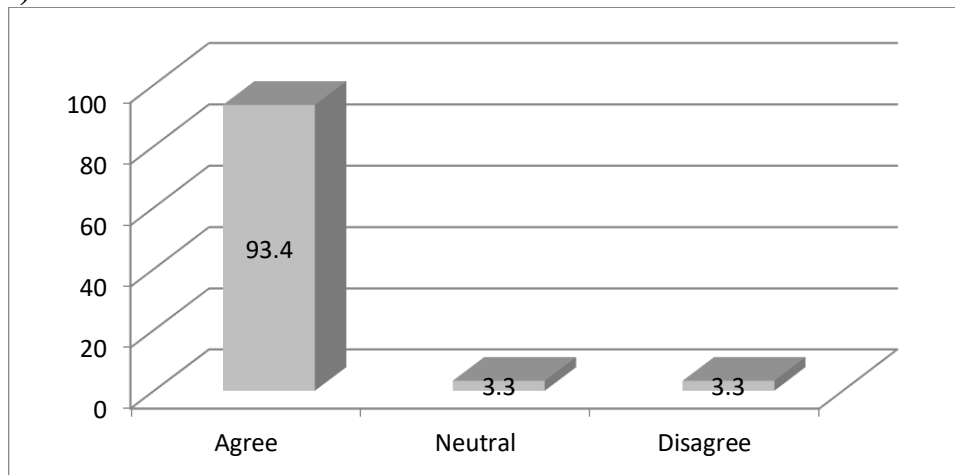
B. Textbooks and Materials

Statement No. (1): Teachers follow and communicate the lesson objectives with the students while planning their lessons.

Table No (4.15) The Frequency Distribution for the Respondents' Answers of Statement No (1)

Valid	Frequency	Percentage
Agree	28	93.4
Neutral	1	3.3
Disagree	1	3.3
Total	30	100%

Figure (4.9)



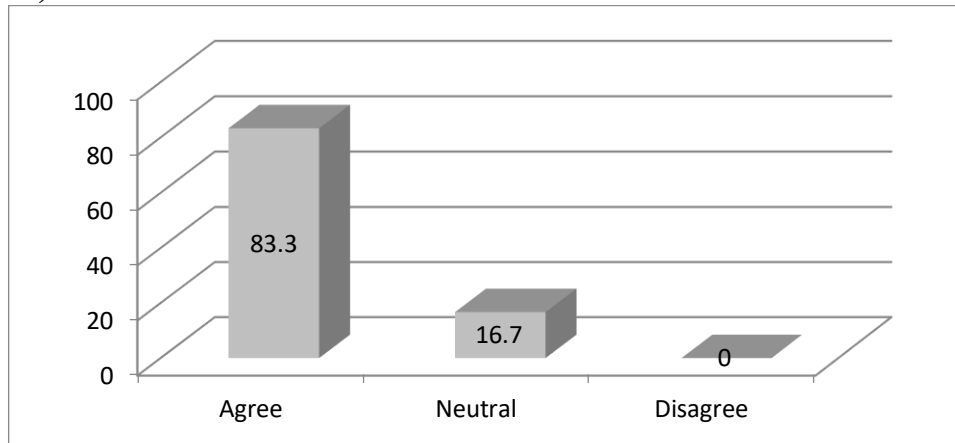
It is obvious from the above table and figure that there are (28) respondents in the study sample with percentage (93.4%) agreed with that “Teachers follow and communicate the lesson objectives with the students while planning their lessons. ”. and (1) respondents with percentage (3.3%) were neutral and (1) respondents with percentage (3.3%) disagreed.

Statement No. (2) The textbook (*Traveller*) covers exercises that the curriculum has claimed

Table No (4.16) The Frequency Distribution for the Respondents' Answers of Statement No (2)

Valid	Frequency	Percentage
Agree	25	83.3
Neutral	5	16.7
Disagree	0	0
Total	30	100%

Figure (4.10)



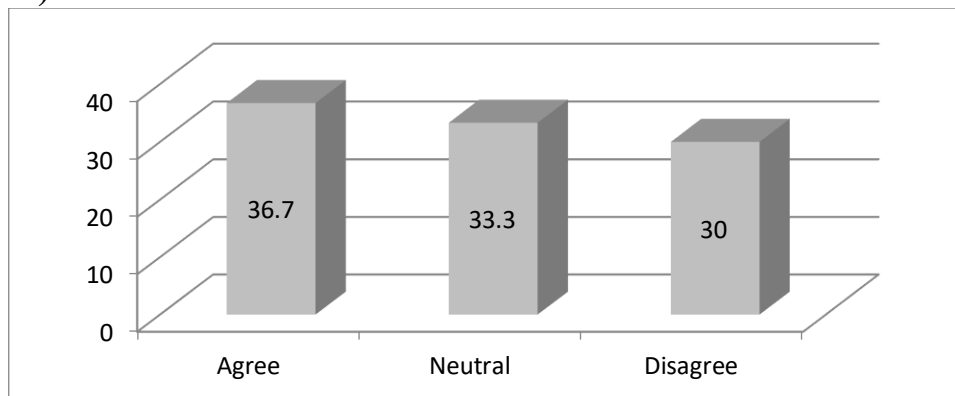
It is obvious from the above table and figure that there are (25) respondents in the study sample with percentage (83.3%) agreed with that "The textbook (*Traveller*) covers exercises that the curriculum has claimed.". and (5) respondents with percentage (16.7%) were neutral and (0) respondents with percentage (0%) disagreed.

Statement No. (3): Teachers think their students do not seriously study the textbook materials.

Table No (4.17) The Frequency Distribution for the Respondents' Answers of Statement No (3)

Valid	Frequency	Percentage
Agree	11	36.7
Neutral	10	33.3
Disagree	9	30
Total	30	100%

Figure (4.11)



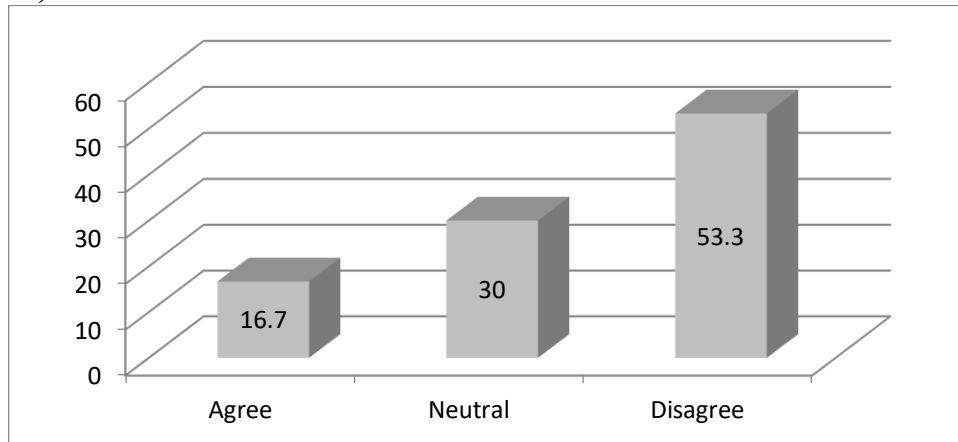
It is obvious from the above table and figure that there are (11) respondents in the study sample with percentage (36.7%) agreed with that "Teachers think their students do not seriously study the textbook materials. ". and (10) respondents with percentage (33.3%) were neutral and (9) respondents with percentage (30.0%) disagreed.

Statement No. (4): Teachers skip certain topics in the textbook because they are less likely to be tested in the examination.

Table No (4.18) The Frequency Distribution for the Respondents' Answers of Statement No (4)

Valid	Frequency	Percentage
Agree	5	16.7
Neutral	9	30
Disagree	16	53.3
Total	30	100%

Figure (4.12)



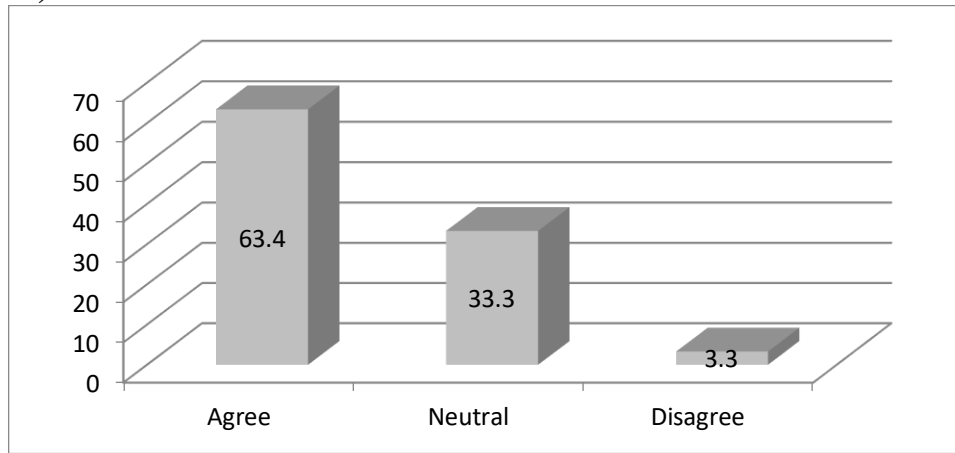
It is obvious from the above table and figure that there are (5) respondents in the study sample with percentage (16.7%) agreed with that “Teachers skip certain topics in the textbook because they are less likely to be tested in the examination. ”. and (9) respondents with percentage (30%) were neutral and (16) respondents with percentage (53.3%) disagreed.

Statement No. (5):The textbook (*Traveller*) is well-suited to developing the students' communicative competence.

Table No (4.19) The Frequency Distribution for the Respondents' Answers of Statement No (5)

Valid	Frequency	Percentage
Agree	19	63.4
Neutral	10	33.3
Disagree	1	3.3
Total	30	100%

Figure (4.13)



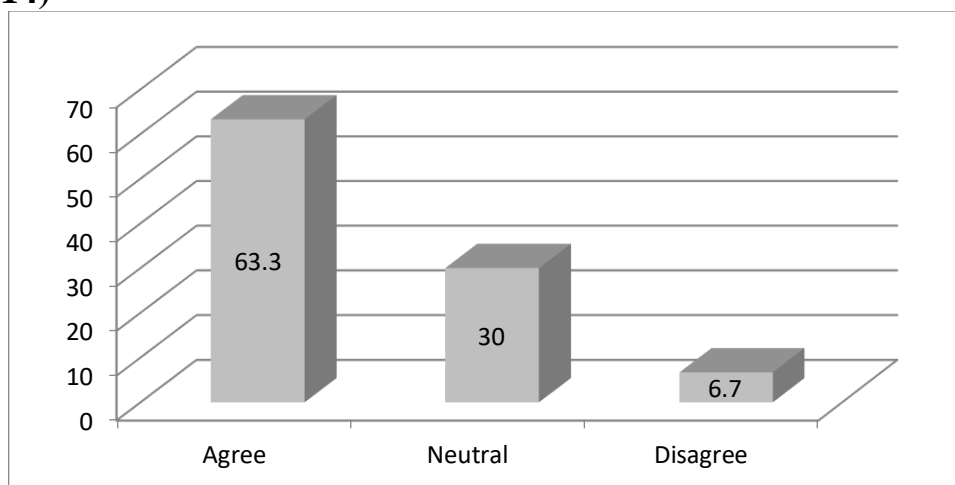
It is obvious from the above table and figure that there are (19) respondents in the study sample with percentage (63.4%) agreed with that “The textbook (*Traveller*) is well-suited to developing the students’ communicative competence. ”. and (10) respondents with percentage (33.3%) were neutral and (1) respondents with percentage (3.3%) disagreed.

Statement No. (6): Teachers rely on test papers and sample questions to prepare their students for the final test.

Table No (4.20) The Frequency Distribution for the Respondents’ Answers of **Statement No (6)**

Valid	Frequency	Percentage
Agree	19	63.3
Neutral	9	30
Disagree	2	6.7
Total	30	100%

Figure (4.14)



It is obvious from the above table and figure that there are (19) respondents in the study sample with percentage (63.3%) agreed with that "Teachers rely on test

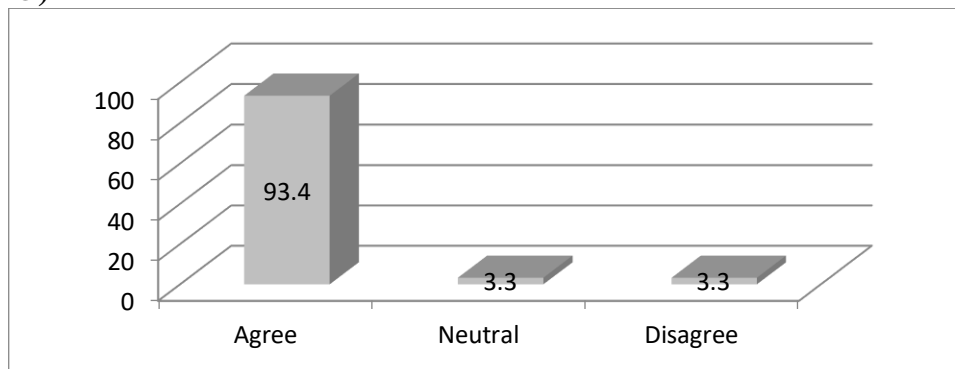
papers and sample questions to prepare their students for the final test. "and (9) respondents with percentage (30%) were neutral and (2) respondents with percentage (6.7%) disagreed.

Statement No. (7): Teachers find interest in teaching textbook materials.

Table No (4.21) The Frequency Distribution for the Respondents' Answers of Statement No (7)

Valid	Frequency	Percentage
Agree	28	93.4
Neutral	1	3.3
Disagree	1	3.3
Total	30	100%

Figure (4.15)



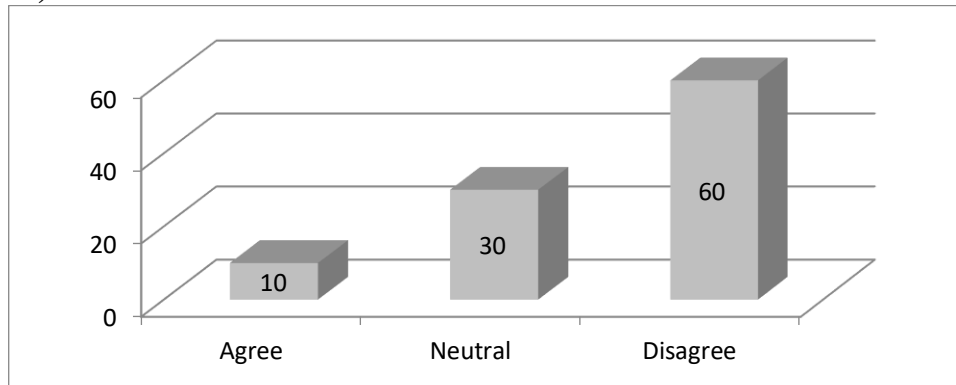
It is obvious from the above table and figure that there are (28) respondents in the study sample with percentage (93.4%) agreed with that “Teachers find interest in teaching textbook materials. ”. and (1) respondents with percentage (3.3%) were neutral and (1) respondents with percentage (3.3%) disagreed.

Statement No. (8): If teachers teach the whole textbook (*Traveller*), their students will perform badly in the test.):

Table No (4.22) The Frequency Distribution for the Respondents' Answers of Statement No (8)

Valid	Frequency	Percentage
Agree	3	10
Neutral	9	30
Disagree	18	60
Total	30	100%

Figure (4.16)



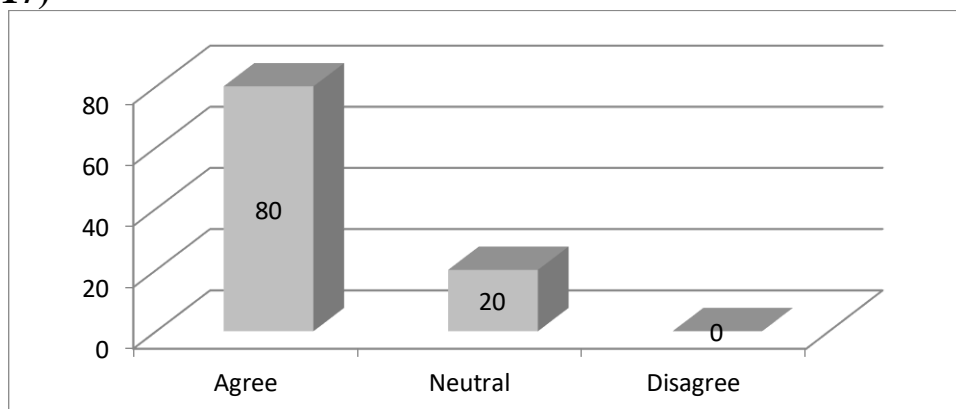
It is obvious from the above table and figure that there are (3) respondents in the study sample with percentage (10.0%) agreed with that “If Teachers teach the whole textbook (*Traveller*), their students will perform badly in the test. ”. and (9) respondents with percentage (30%) were neutral and (18) respondents with percentage (60.0%) disagreed.

Statement No. (9): Teachers use authentic materials along with the textbook for the students’ practice of English language in the class.

Table No (4.23) The Frequency Distribution for the Respondents’ Answers of Statement No (9)

Valid	Frequency	Percentage
Agree	24	80
Neutral	6	20
Disagree	0	0
Total	30	100%

Figure (4.17)



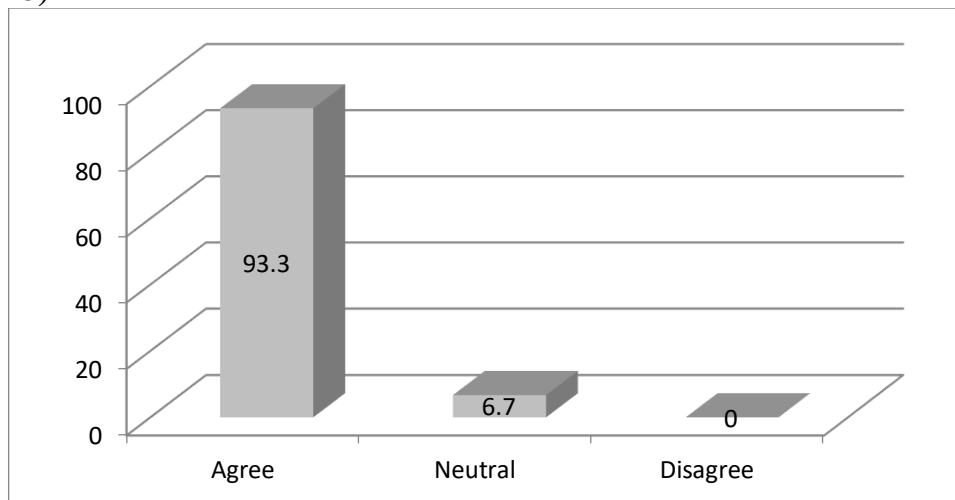
It is obvious from the above table and figure that there are (24) respondents in the study sample with percentage (80.0%) agreed with that "Teachers use authentic materials along with the textbook for the students' practice of English language in the class. ". and (6) respondents with percentage (20%) were neutral and (0) respondents with percentage (0.0%) disagreed.

Statement No. (10): Teachers use modern aids and equipment to teach and practice English language skills.

Table No (4.24) The Frequency Distribution for the Respondents' Answers of **Statement No (10)**

Valid	Frequency	Percentage
Agree	28	93.3
Neutral	2	6.7
Disagree	0	0
Total	30	100%

Figure (4.18)



It is obvious from the above table and figure that there are (28) respondents in the study sample with percentage (93.3%) agreed with that "Teachers use modern aids and equipment to teach and practice English language skills." and (2) respondents with percentage (6.7%) were neutral and (0) respondents with percentage (0.0%) disagreed.

C. Teaching Methods

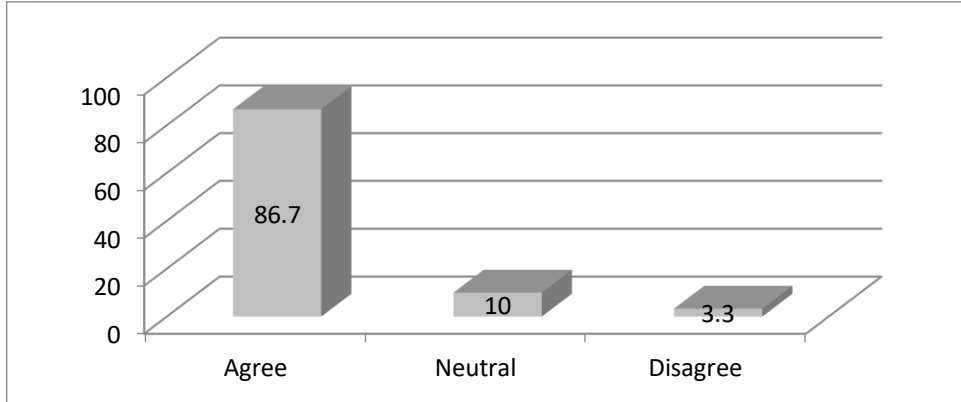
Statement No. (1): Teachers teach in a way that their students understand everything.

Table No (4.25) The Frequency Distribution for the Respondents' Answers of **Statement No (1)**

Valid	Frequency	Percentage
Agree	26	86.7

Neutral	3	10
Disagree	1	3.3
Total	30	100%

Figure (4.19)



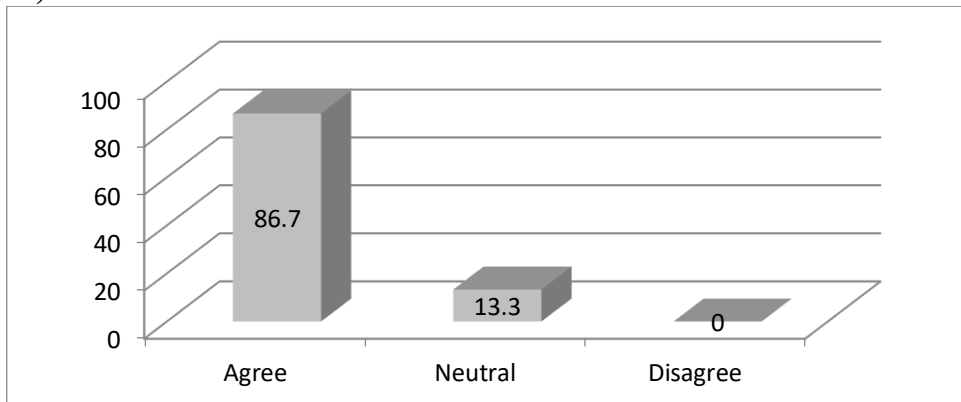
It is obvious from the above table and figure that there are (26) respondents in the study sample with percentage (86.7%) agreed with that “Teachers teach in a way that their students understand everything. ”. and (3) respondents with percentage (10%) were neutral and (1) respondents with percentage (3.3%) disagreed.

Statement No. (2): Teachers teach and explain the text in English.

Table No (4.26) The Frequency Distribution for the Respondents’ Answers of **Statement No (2)**

Valid	Frequency	Percentage
Agree	26	86.7
Neutral	4	13.3
Disagree	0	0
Total	30	100%

Figure (4.20)



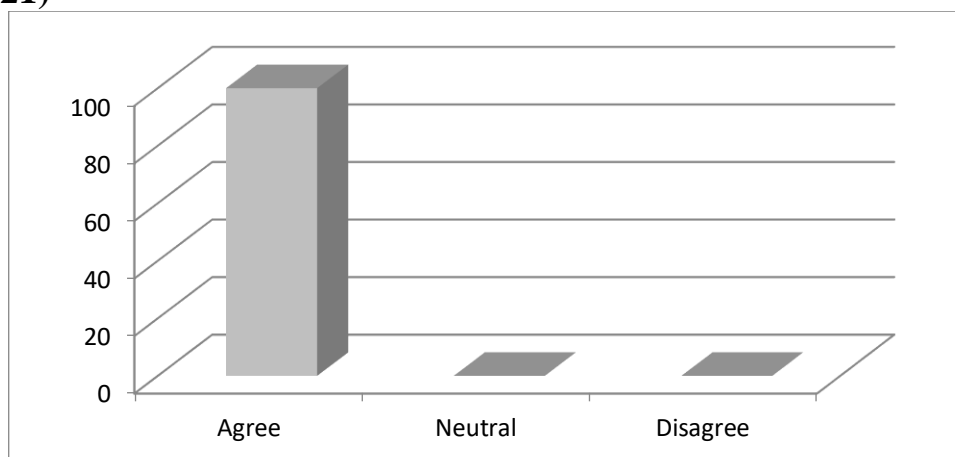
It is obvious from the above table and figure that there are (20) respondents in the study sample with percentage (86.7%) agreed with that “Teachers teach and explain the text in English. ”. and (4) respondents with percentage (13.3%) were neutral and (0) respondents with percentage (0.0%) disagreed.

Statement No. (3): Teachers encourage their students to ask questions during the class.

Table No (4.27) The Frequency Distribution for the Respondents’ Answers of Statement No (3)

Valid	Frequency	Percentage
Agree	30	100
Neutral	0	0
Disagree	0	0
Total	30	100%

Figure (4.21)



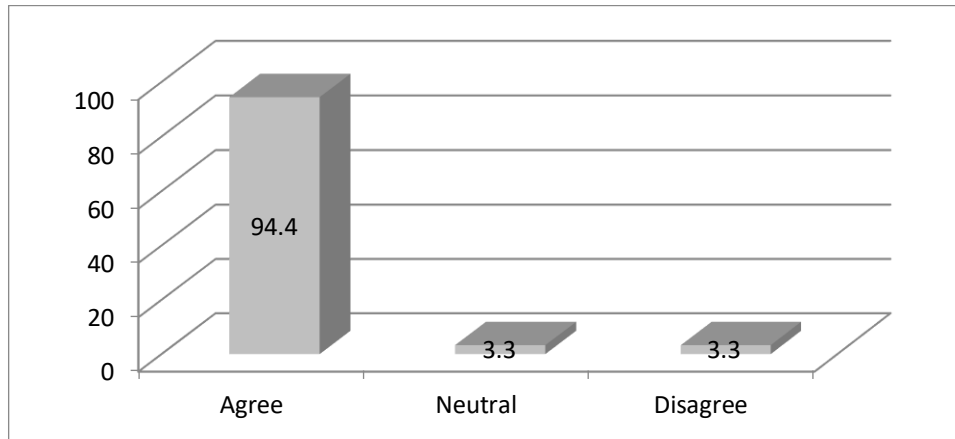
It is obvious from the above table and figure that there are (30) respondents in the study sample with percentage (100.0%) agreed with that “Teachers encourage their students to ask questions during the class. ”. and (0) respondents with percentage (0%) were neutral and (0) respondents with percentage (0%) disagreed.

Statement No. (4): Teachers encourage their students to speak English in the class.

Table No (4.28) The Frequency Distribution for the Respondents’ Answers of Statement No (4)

Valid	Frequency	Percentage
Agree	28	94.4
Neutral	1	3.3
Disagree	1	3.3
Total	30	100%

Figure (4.22)



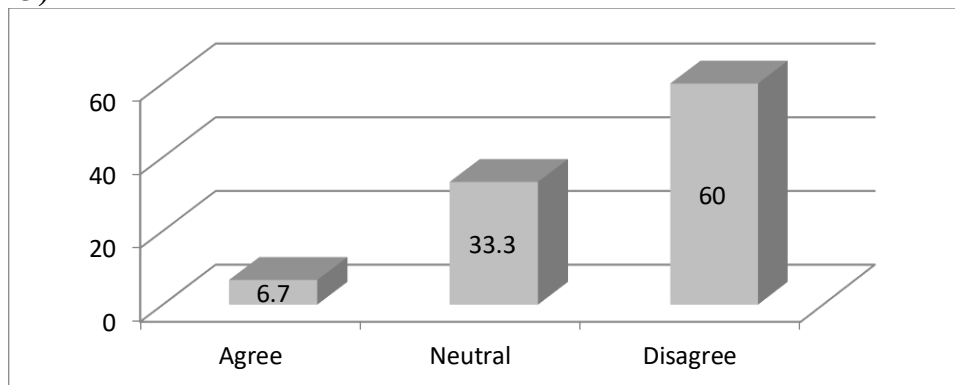
It is obvious from the above table and figure that there are (28) respondents in the study sample with percentage (94.4%) agreed with that “Teachers encourage their students to speak English in the class. ”. and (1) respondents with percentage (3.3%) were neutral and (1) respondents with percentage (3.3%) disagreed.

Statement No. (5): Teachers use Arabic along with English to make their students understand better.

Table No (4.29) The Frequency Distribution for the Respondents’ Answers of **Statement No (5)**

Valid	Frequency	Percentage
Agree	2	6.7
Neutral	10	33.3
Disagree	18	60
Total	30	100%

Figure (4.23)



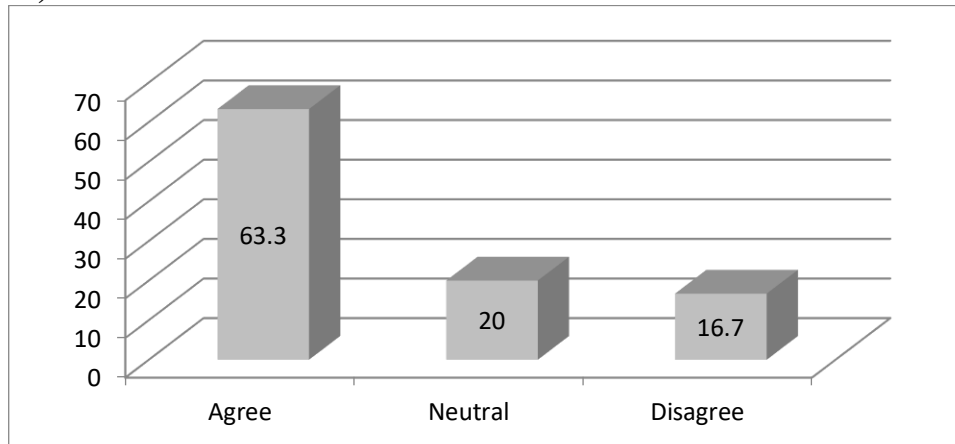
It is obvious from the above table and figure that there are (2) respondents in the study sample with percentage (6.7%) agreed with that "Teachers use Arabic along with English to make their students understand better. ". and (10) respondents with percentage (33.3%) were neutral and (18) respondents with percentage (60.0%) disagreed.

Statement No. (6): Teachers teach whatever they think important to teach, no matter whether it is important or not for the test.

Table No (4.30) The Frequency Distribution for the Respondents' Answers of Statement No (6)

Valid	Frequency	Percentage
Agree	19	63.3
Neutral	6	20
Disagree	5	16.7
Total	30	100%

Figure (4.24)



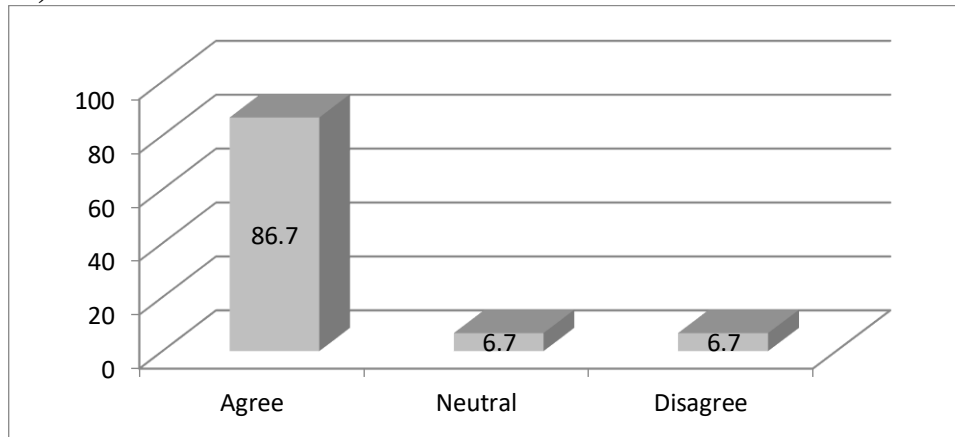
It is obvious from the above table and figure that there are (19) respondents in the study sample with percentage (63.3%) agreed with that " Teachers teach whatever they think important to teach, no matter whether it is important or not for the test. ". and (6) respondents with percentage (20%) were neutral and (5) respondents with percentage (16.7%) disagreed.

Statement No. (7): The role of teachers of English is to transmit knowledge to their students through explaining texts and giving examples.

Table No (4.31) The Frequency Distribution for the Respondents' Answers of Statement No (7)

Valid	Frequency	Percentage
Agree	26	86.7
Neutral	2	6.7
Disagree	2	6.7
Total	30	100%

Figure (4.25)



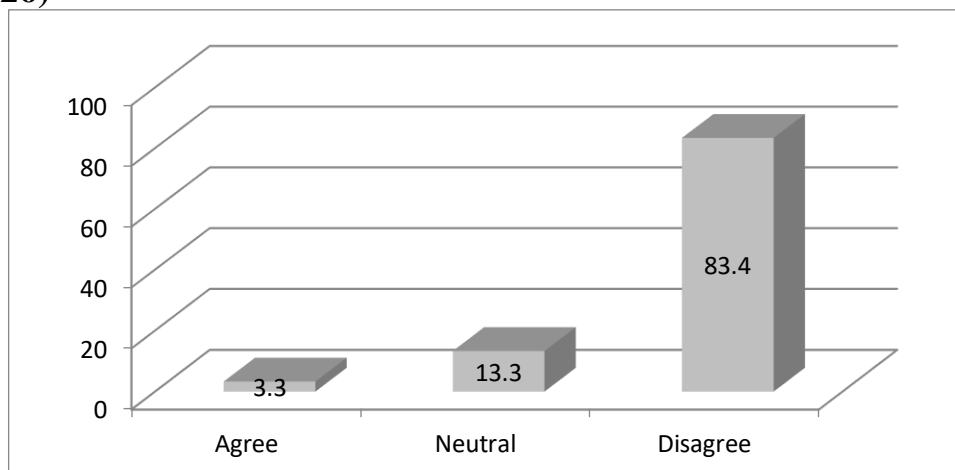
It is obvious from the above table and figure that there are (26) respondents in the study sample with percentage (86.7%) agreed with that “The role teachers of English is to transmit knowledge to their students through explaining texts and giving examples. ”. and (2) respondents with percentage (6.7%) were neutral and (2) respondents with percentage (6.7%) disagreed.

Statement No. (8): Teachers do not make their students practice how to learn and speak English language but make them practice on how to answer questions in the test.

Table No (4.32) (4.31) The Frequency Distribution for the Respondents’ Answers of Statement No (8)

Valid	Frequency	Percentage
Agree	1	3.3
Neutral	4	13.3
Disagree	25	83.4
Total	30	100%

Figure (4.26)



It is obvious from the above table and figure that there are (1) respondents in the study sample with percentage (3.3%) agreed with that " Teachers do not make their

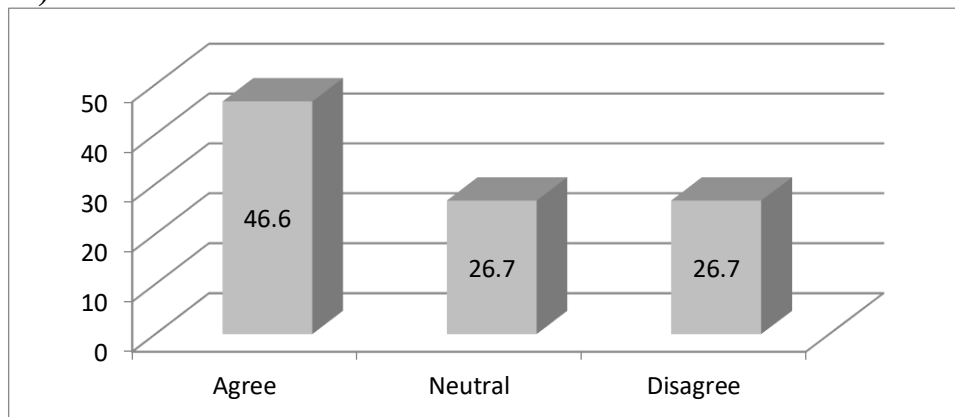
students practice how to learn and speak English language but make them practice on how to answer questions in the test. ". and (4) respondents with percentage (13.3%) were neutral and (25) respondents with percentage (83.4%) disagreed.

Statement No. (9): Teachers believe that the test score in English is an appropriate indicator of a student's English ability.

Table No (4.33) The Frequency Distribution for the Respondents' Answers of Statement No (9)

Valid	Frequency	Percentage
Agree	14	46.6
Neutral	8	26.7
Disagree	8	26.7
Total	30	100%

Figure (4.27)



It is obvious from the above table and figure that there are (14) respondents in the study sample with percentage (46.6%) agreed with that “Teachers believe that the test score in English is an appropriate indicator of a student's English ability. ”. and (8) respondents with percentage (26.7%) were neutral and (8) respondents with percentage (26. 7%) disagreed.

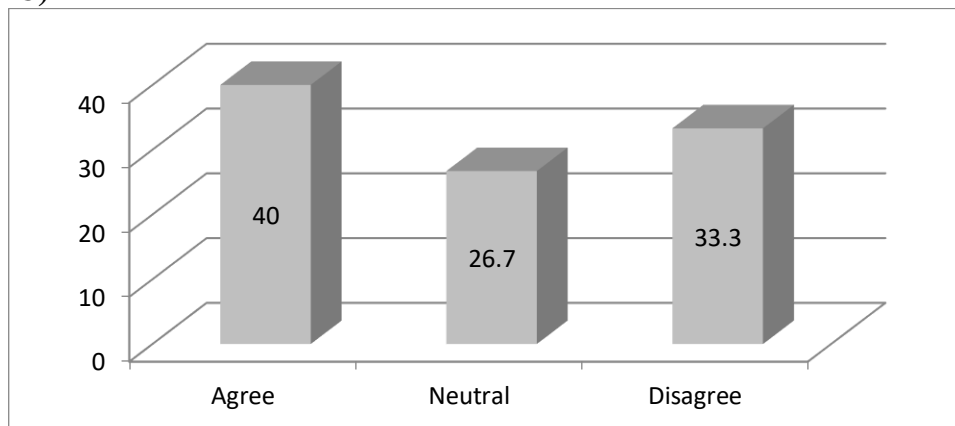
D. Classroom Tasks and Activities

Statement No. (1): Teachers ignore the task and activities that are not directly related to passing the test.

Table No (4.34) The Frequency Distribution for the Respondents' Answers of Statement No (1)

Valid	Frequency	Percentage
Agree	12	40
Neutral	8	26.7
Disagree	10	33.3
Total	30	100%

Figure (4.28)



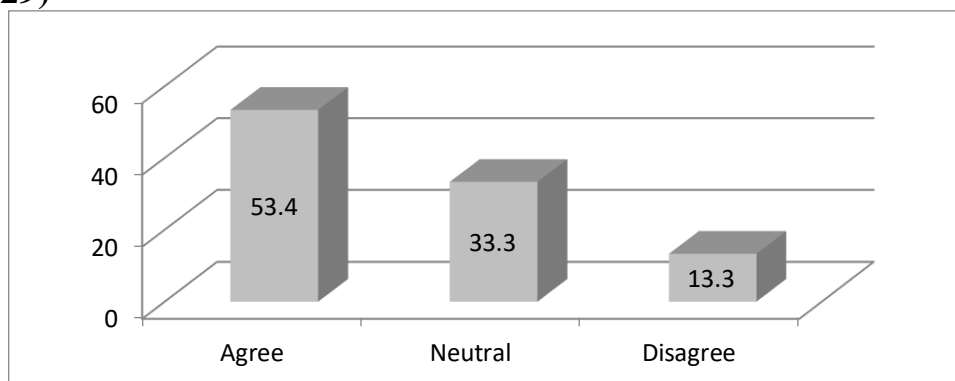
It is obvious from the above table and figure that there are (12) respondents in the study sample with percentage (40.0%) agreed with that “Teachers ignore the task and activities that are not directly related to passing the test. ”. and (8) respondents with percentage (26.7%) were neutral and (10) respondents with percentage (33.3%) disagreed.

Statement No. (2): Teachers give model tests to the students to do better in the final test.

Table No (4.35) The Frequency Distribution for the Respondents’ Answers of Statement No (2)

Valid	Frequency	Percentage
Agree	16	53.4
Neutral	10	33.3
Disagree	4	13.3
Total	30	100%

Figure (4.29)



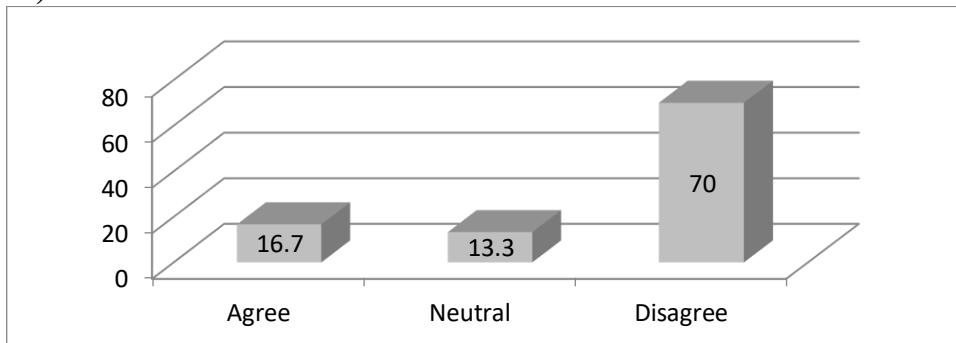
It is obvious from the above table and figure that there are (16) respondents in the study sample with percentage (53.4%) agreed with that “teachers give model tests to the students to do better in the final test. ”. and (10) respondents with percentage (33.3%) were neutral and (4) respondents with percentage (13.3%) disagreed.

Statement No. (3): Teachers spend more time teaching grammar because they think grammar is more likely to be tested.

Table No (4.36) The Frequency Distribution for the Respondents' Answers of Statement No (3)

Valid	Frequency	Percentage
Agree	5	16.7
Neutral	4	13.3
Disagree	21	70
Total	30	100%

Figure (4.30)



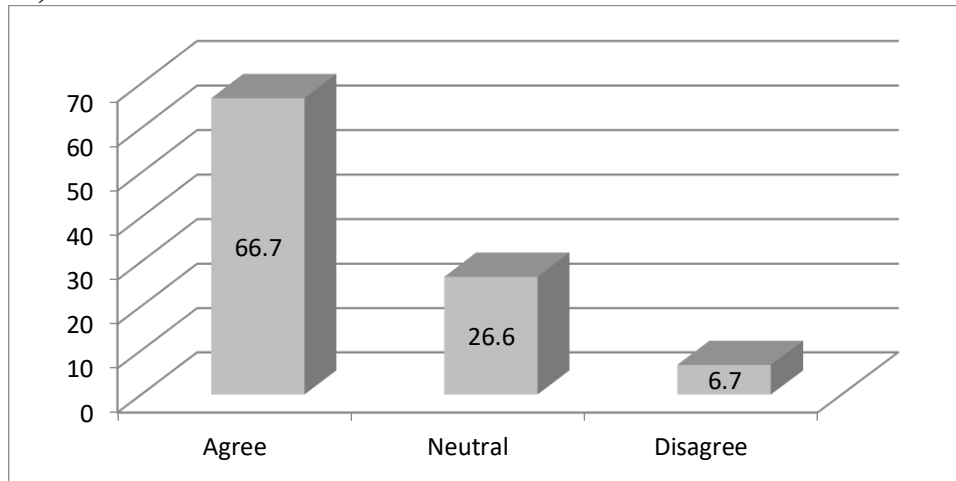
It is obvious from the above table and figure that there are (5) respondents in the study sample with percentage (16.7%) agreed with that “teachers spend more time teaching grammar because they think grammar is more likely to be tested. ”. and (4) respondents with percentage (13.3%) were neutral and (21) respondents with percentage (70.0%) disagreed.

Statement No. (4): Teachers make their students practice and solve the questions of the past tests.

Table No (4.37) The Frequency Distribution for the Respondents' Answers of Statement No (4)

Valid	Frequency	Percentage
Agree	20	66.7
Neutral	8	26.6
Disagree	2	6.7
Total	30	100%

Figure (4.31)



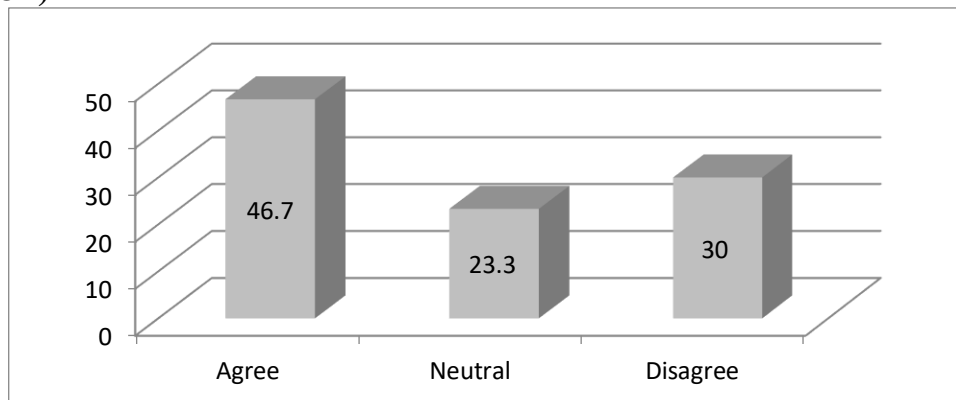
It is obvious from the above table and figure that there are (20) respondents in the study sample with percentage (66.7%) agreed with that "Teachers make their students practice and solve the questions of the past tests. ", and (8) respondents with percentage (26.6%) were neutral and (2) respondents with percentage (6.7%) disagreed.

Statement No. (5): The test discourages the teachers to teach English language

Table No (4.38) The Frequency Distribution for the Respondents' Answers of Statement No (5)

Valid	Frequency	Percentage
Agree	14	46.7
Neutral	7	23.3
Disagree	9	30
Total	30	100%

Figure (4.32)



It is obvious from the above table and figure that there are (14) respondents in the study sample with percentage (46.7%) agreed with that " The test discourages the

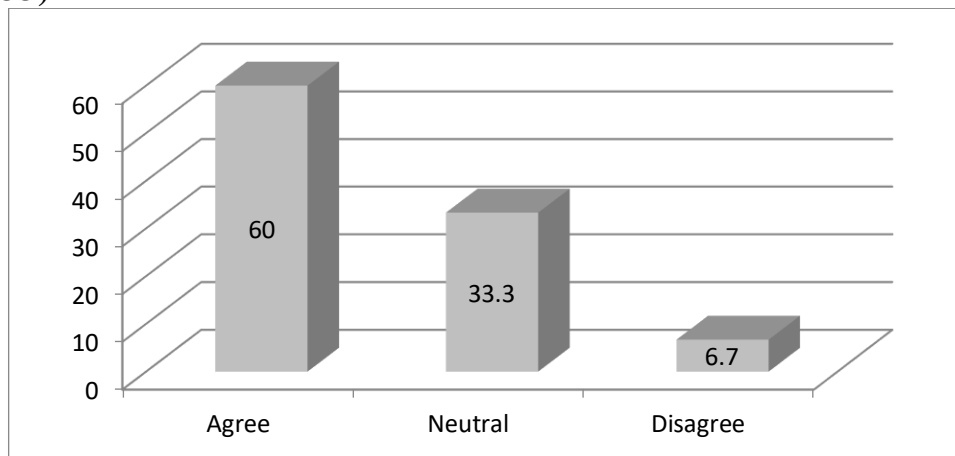
teachers to teach English language. ". and (7) respondents with percentage (23.3%) were neutral and (9) respondents with percentage (30.0%) disagreed.

Statement No. (6): Teachers teach test-taking strategies, especially when the test date gets closer.

Table No (4.39) The Frequency Distribution for the Respondents' Answers of Statement No (6)

Valid	Frequency	Percentage
Agree	18	60
Neutral	10	33.3
Disagree	2	6.7
Total	30	100%

Figure (4.33)



It is obvious from the above table and figure that there are (18) respondents in the study sample with percentage (60.0%) agreed with that " Teachers teach test-taking strategies, especially when the test date gets closer. ". and (10) respondents with percentage (33.3%) were neutral and (2) respondents with percentage (6.7.3%) disagreed.

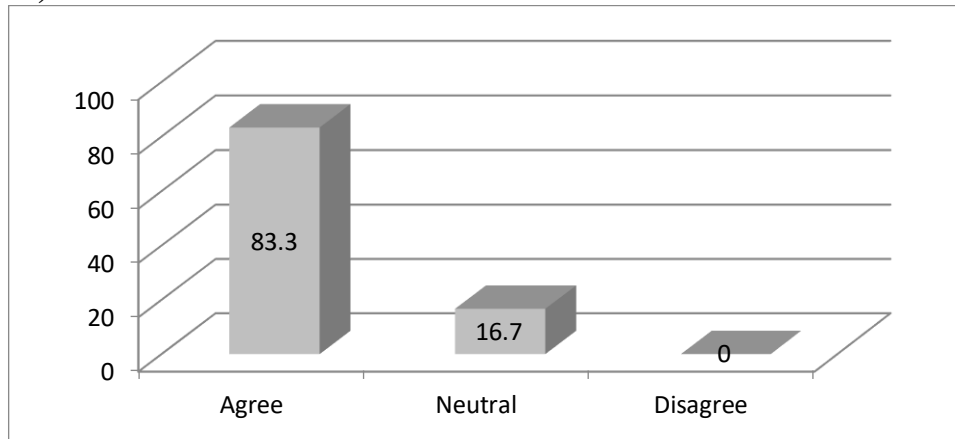
E. EFL Skills and Elements

Statement No. (1): The test influences the teacher's decision on which language skills are more important to be taught.

Table No (4.40) The Frequency Distribution for the Respondents' Answers of Statement No (1)

Valid	Frequency	Percentage
Agree	25	83.3
Neutral	5	16.7
Disagree	0	0
Total	30	100%

Figure (4.34)



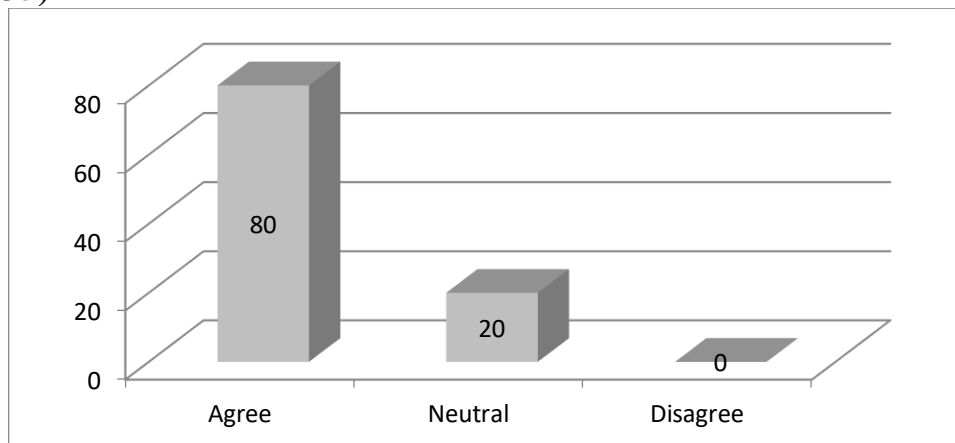
It is obvious from the above table and figure that there are (11) respondents in the study sample with percentage (83.3%) agreed with that "The test influences the teacher's decision on which language skills are more important to be taught." and (12) respondents with percentage (16.7%) were neutral and (7) respondents with percentage (0.0%) disagreed.

Statement No. (2): Listening is practiced in the class.

Table No (4.41) The Frequency Distribution for the Respondents' Answers of Statement No (2)

Valid	Frequency	Percentage
Agree	24	80
Neutral	6	20
Disagree	0	0
Total	30	100%

Figure (4.35)



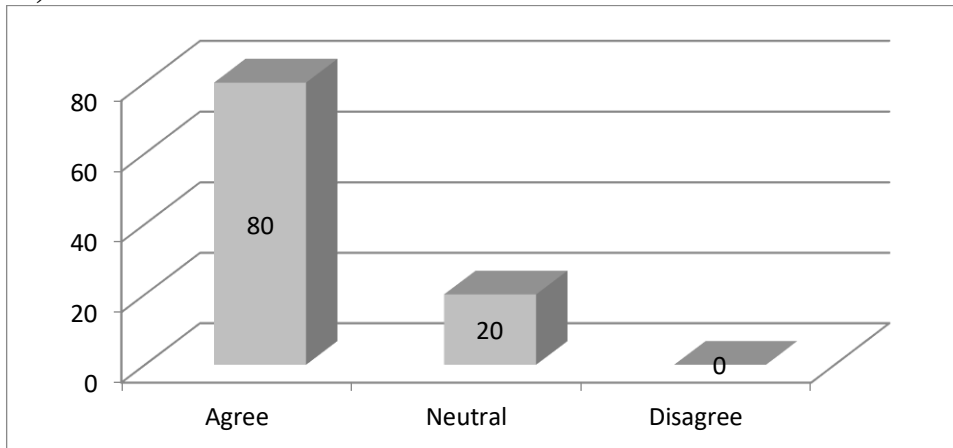
It is obvious from the above table and figure that there are (24) respondents in the study sample with percentage (80.0%) agreed with that "Listening is practiced in the class.". and (6) respondents with percentage (20%) were neutral and (0) respondents with percentage (0.0%) disagreed.

Statement No. (3): Speaking is practiced in the class.

Table No (4.42) The Frequency Distribution for the Respondents' Answers of Statement No (3)

Valid	Frequency	Percentage
Agree	24	80
Neutral	6	20
Disagree	0	0
Total	30	100%

Figure (4.36)



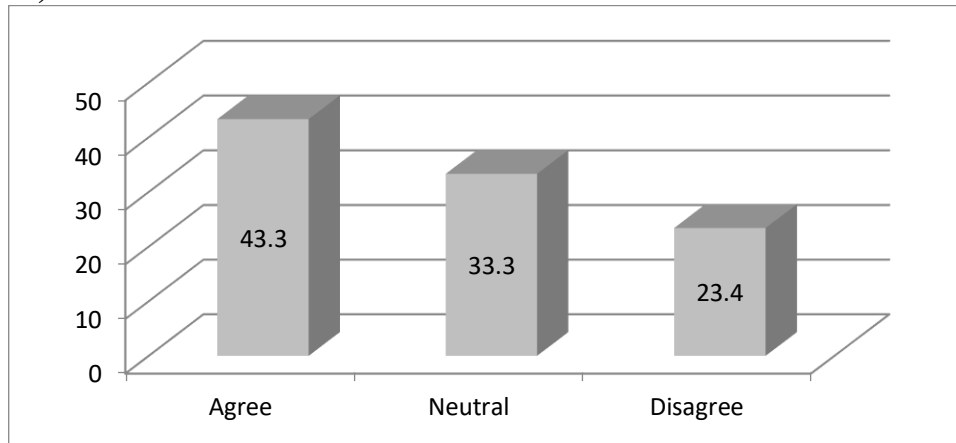
It is obvious from the above table and figure that there are (24) respondents in the study sample with percentage (80.0%) agreed with that "Speaking is practiced in the class.. ". and (6) respondents with percentage (20%) were neutral and (0) respondents with percentage (0.0%) disagreed.

Statement No. (4): Reading is practiced in the class.

Table No (4.43) The Frequency Distribution for the Respondents' Answers of Statement No (4)

Valid	Frequency	Percentage
Agree	13	43.3
Neutral	10	33.3
Disagree	7	23.4
Total	30	100%

Figure (4.37)



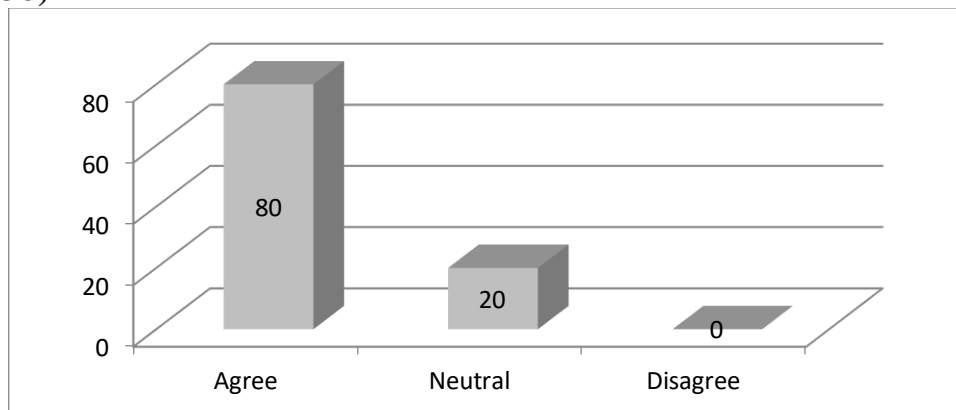
It is obvious from the above table and figure that there are (13) respondents in the study sample with percentage (43.3%) agreed with that “Reading is practiced in the class. ”. and (10) respondents with percentage (33.3%) were neutral and (7) respondents with percentage (23.4%) disagreed.

Statement No. (5): Writing is practiced in the class.

Table No (4.44) The Frequency Distribution for the Respondents’ Answers of Statement No (5)

Valid	Frequency	Percentage
Agree	24	80
Neutral	6	20
Disagree	0	0
Total	30	100%

Figure (4.38)



It is obvious from the above table and figure that there are (24) respondents in the study sample with percentage (80.0%) agreed with that “Writing is practiced in the class. ”. and (6) respondents with percentage (20%) were neutral and (0) respondents with percentage (0.0%) disagreed.

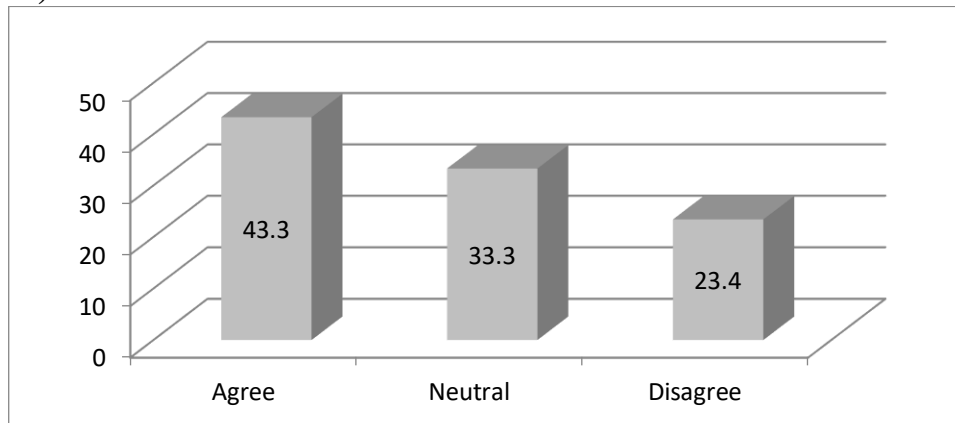
F. Beliefs, Attitudes, and Perception Related to the Test and Teaching

Statement No. (1): Teachers feel pressure from their authority to improve their students' test score.

Table No (4.45) The Frequency Distribution for the Respondents' Answers of Statement No (1)

Valid	Frequency	Percentage
Agree	13	43.3
Neutral	10	33.3
Disagree	7	23.4
Total	30	100%

Figure (4.39)



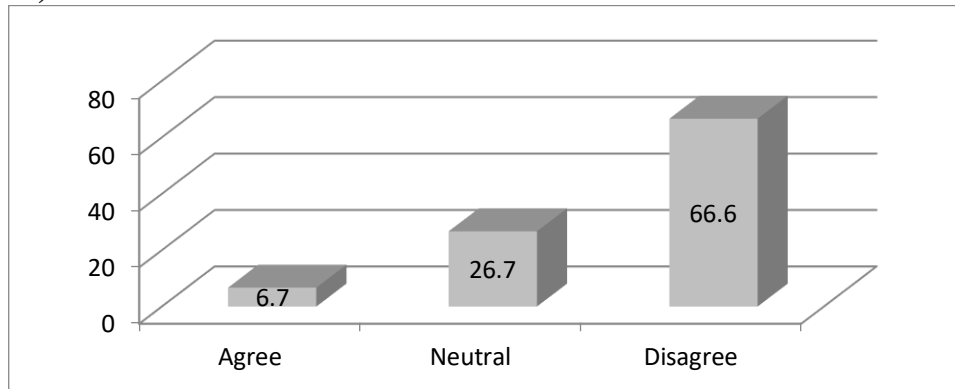
It is obvious from the above table and figure that there are (13) respondents in the study sample with percentage (43.3%) agreed with that “Teachers feel pressure from their authority to improve their students' test score. ”. and (10) respondents with percentage (33.3%) were neutral and (7) respondents with percentage (23.4%) disagreed.

Statement No. (2): Students can score good marks without improving their English language proficiency.

Table No (4.46) The Frequency Distribution for the Respondents' Answers of Statement No (2)

Valid	Frequency	Percentage
Agree	2	6.7
Neutral	8	26.7
Disagree	20	66.6
Total	30	100%

Figure (4.40)



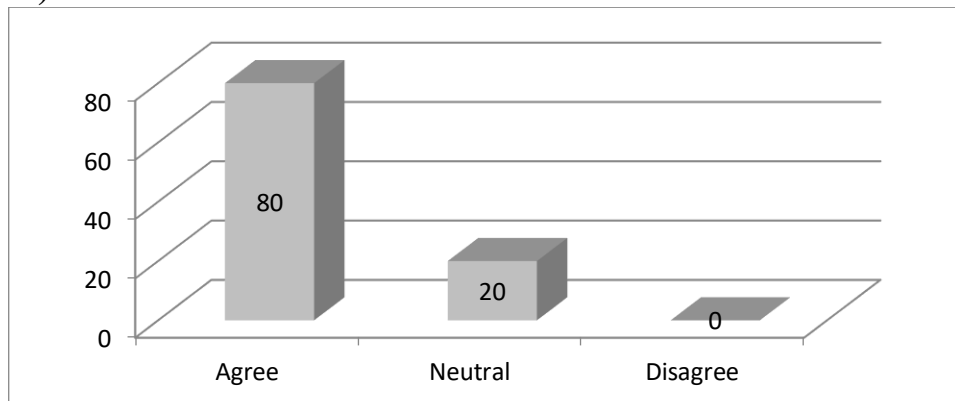
It is obvious from the above table and figure that there are (2) respondents in the study sample with percentage (6.7%) agreed with that " Students can score good marks without improving their English language proficiency.. ". and (8) respondents with percentage (26.7%) were neutral and (20) respondents with percentage (66.6%) disagreed.

Statement No. (3): Teachers get feedback on their teaching from the test results.

Table No (4.47) The Frequency Distribution for the Respondents' Answers of Statement No (3)

Valid	Frequency	Percentage
Agree	24	80
Neutral	6	20
Disagree	0	0
Total	30	100%

Figure (4.41)



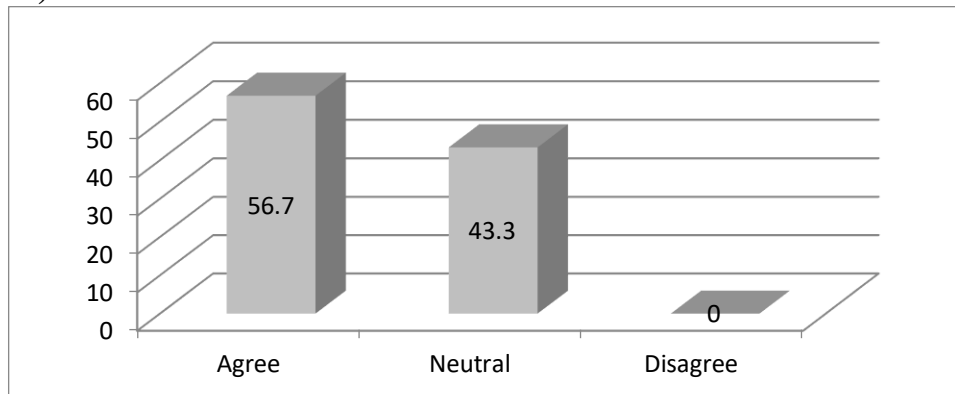
It is obvious from the above table and figure that there are (24) respondents in the study sample with percentage (80.0%) agreed with that "Teachers get feedback on their teaching from the test results. ". and (6) respondents with percentage (20%) were neutral and (0) respondents with percentage (60.0%) disagreed.

Statement No. (4): Students suffer from anxiety and tension for the test.

Table No (4.48) The Frequency Distribution for the Respondents' Answers of Statement No (4)

Valid	Frequency	Percentage
Agree	17	56.7
Neutral	13	43.3
Disagree	0	0
Total	30	100%

Figure (4.42)



It is obvious from the above table and figure that there are (17) respondents in the study sample with percentage (56.7%) agreed with that "Students suffer from anxiety and tension for the test.

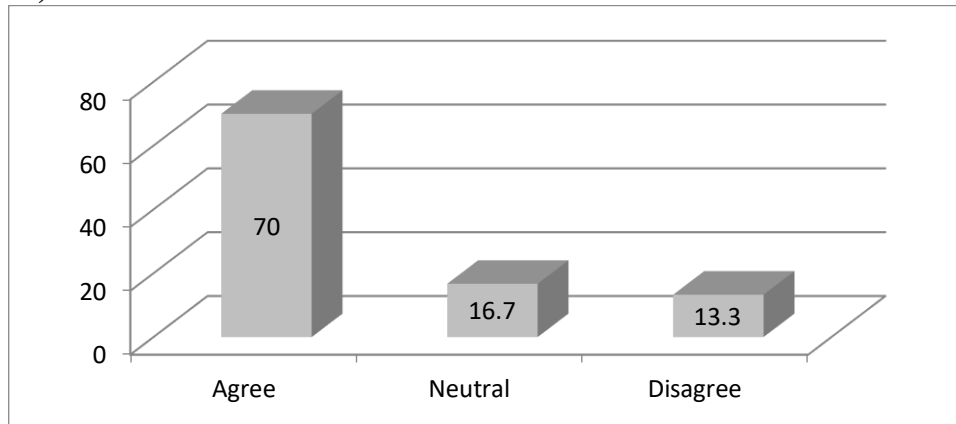
.". and (13) respondents with percentage (43.3%) were neutral and (0) respondents with percentage (0.0%) disagreed.

Statement No. (5): The present testing system helps the students improve language proficiency.

Table No (4.49) The Frequency Distribution for the Respondents' Answers of Statement No (5)

Valid	Frequency	Percentage
Agree	21	70
Neutral	5	16.7
Disagree	4	13.3
Total	30	100%

Figure (4.43)



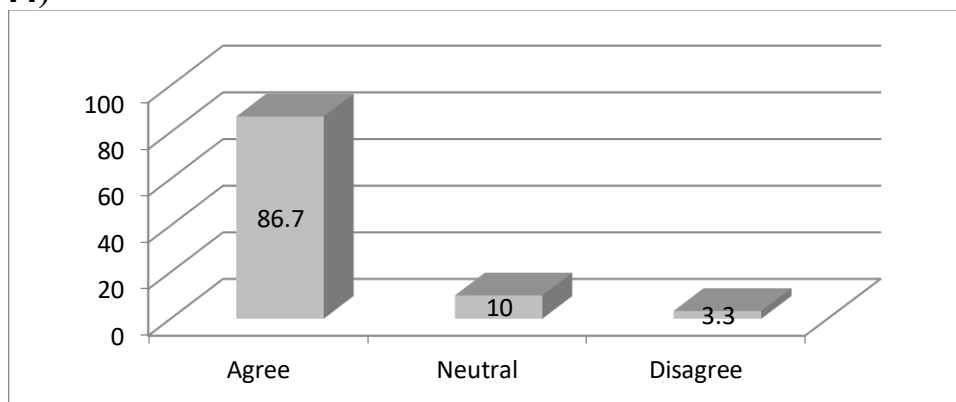
It is obvious from the above table and figure that there are (21) respondents in the study sample with percentage (70.0%) agreed with that “The present testing system helps the students improve language proficiency.”. and (5) respondents with percentage (16.7%) were neutral and (4) respondents with percentage (13.3%) disagreed.

Statement No. (6): Teachers could teach English better if there were no pressures for good results in the test.

Table No (4.50) The Frequency Distribution for the Respondents’ Answers of Statement No (6)

Valid	Frequency	Percentage
Agree	26	86.7
Neutral	3	10
Disagree	1	3.3
Total	30	100%

Figure (4.44)



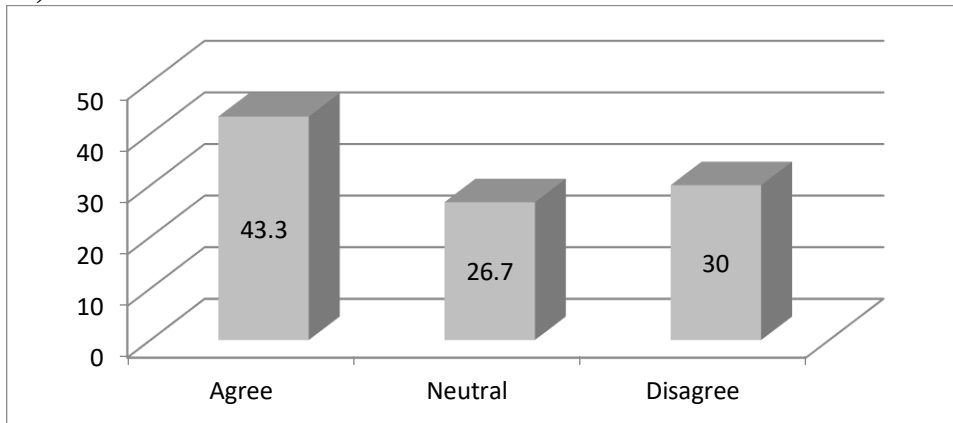
It is obvious from the above table and figure that there are (26) respondents in the study sample with percentage (86.7%) agreed with that "Teachers could teach English better if there were no pressures for good results in the test.. ". and (3) respondents with percentage (10%) were neutral and (1) respondents with percentage (3.3%) disagreed.

Statement No. (7): Tests influence students’ future career.

Table No (4.51) The Frequency Distribution for the Respondents’ Answers of Statement No (7)

Valid	Frequency	Percentage
Agree	13	43.3
Neutral	8	26.7
Disagree	9	30
Total	30	100%

Figure (4.45)



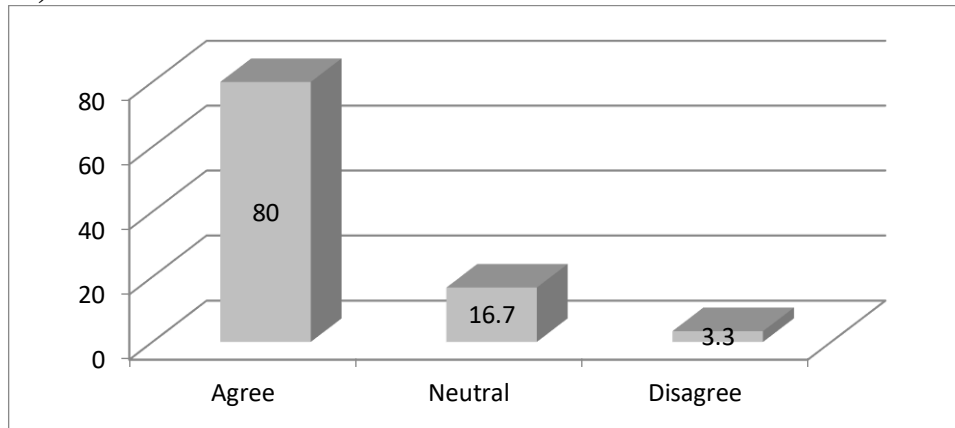
It is obvious from the above table and figure that there are (13) respondents in the study sample with percentage (43.3%) agreed with that “Tests influence students’ future career. ”. and (8) respondents with percentage (26.7%) were neutral and (9) respondents with percentage (30.0%) disagreed.

Statement No. (8): Teachers feel embarrassed if their students fail or perform badly in the test.

Table No (4.52) The Frequency Distribution for the Respondents’ Answers of Statement No (8)

Valid	Frequency	Percentage
Agree	24	80
Neutral	5	16.7
Disagree	1	3.3
Total	30	100%

Figure (4.46)



It is obvious from the above table and figure that there are (24) respondents in the study sample with percentage (80.0%) agreed with that "Teachers feel embarrassed if their students fail or perform badly in the test. ". and (5) respondents with percentage (16.7%) were neutral and (1) respondents with percentage (3.3%) disagreed.

4.5. Testing the Study Hypotheses

Following the analysis and discussion of teachers' questionnaires is devoted for answering the study questions and check its hypotheses that stated in chapter one. The median will be computed for each statement of the questionnaire that shows the opinions of the study respondents about the "washback-based approach of the achievement test in assessing the learning process". To do that, we will give one degree for each answer " agree", "neutral "and "disagree", This means, in accordance with the statistical analysis requirements, transformation of nominal variables to quantitative variables. After that, we will use the non-parametric chi-square test to know if there are statistical differences amongst the respondents' answers about hypotheses statements.

4.5.1 Results of the First Hypothesis:

The First Hypothesis in this study states the following:

“Washback-based approach of the achievement test has positive impact on developing EFL learning process.”

The objective of this hypothesis is to investigate the impact of washback-based approach of the achievement test on developing EFL learning process. To test this hypothesis, we must know the trend of respondents' opinions about each statement

from the hypothesis's statement, and for all statements. We compute the mean, standard deviation, chi square and p-value which is the most central tendency measures, that is used to describe the phenomena, and it represents the centered answer for all respondents' answers after ascending or descending order for the answers.

Table No (4.53) testing the first hypothesis of the study

A. Curriculum and Syllabus

No.	Statement	mean	SD	Chi square	p-value
1.	I am aware of the objectives of the syllabus and curriculum.	2.8	3.4	25	0.000
2.	The present syllabus and curriculum can enhance EFL learning process	2.5	1.5	19	0.000
3.	I teach every section in the textbook (Traveller) although some sections are unlikely to be tested in the examination.	2.4	0.9	31	0.000
4.	I do not care about the syllabus and curriculum while teaching my students.	2.9	1.6	25	0.000
5.	I feel pressure to cover the syllabus before the final test.	2.6	0.7	36	0.000
6.	The test tests the overall competence of my students in English (listening, speaking, reading, and writing).	2.7	1.5	23	0.000
7.	I give more attention to teaching to the syllabus opposed to practicing the test items.).	2.8	0.6	27	0.000

1-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (A-1) was (25) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the

calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (A-1). According to the previous results we can say that our hypothesis is accepted

2-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (A-2) was (19) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). this indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (A-2) According to the previous results we can say that our hypothesis is accepted

3-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (A-3) was (31) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (A-3). According to the previous results we can say that our hypothesis is accepted

4-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (A-4) was (25) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). this indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (A-4) According to the previous results we can say that our hypothesis is accepted

5-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (A-5) was (36) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). this indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (A-5) According to the previous results we can say that our hypothesis is accepted

6-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (A-6) was (23) which is greater than the tabulated

value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). this indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (A-6) According to the previous results we can say that our hypothesis is accepted

7-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (A-7) was (27) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). this indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (A-7) According to the previous results we can say that our hypothesis is accepted

4.5.2 Results of the Second Hypothesis:

The Second Hypothesis in this study states the following:

“Washback-based approach of the achievement test can promote EFL learners' proficiency.”

The objective of this hypothesis is to investigate to what extend washback-based approach of the achievement test can promote EFL learners' proficiency. To test this hypothesis, we must know the trend of respondents' opinions about each statement from the hypothesis's statement, and for all statements. We compute the mean, standard deviation, chi square and p-value which is the most central tendency measures, that is used to describe the phenomena, and it represents the centered answer for all respondents' answers after ascending or descending order for the answers.

Table No (4.54) testing the second hypothesis of the study

B. Textbooks and Materials

No.	Statement	mean	SD	Chi square	p-value
1.	I follow and communicate the lesson objectives with the students while planning my lessons.	2.5	1.5	19	0.000
2.	The textbook (Traveller) covers exercises that the curriculum has claimed.	2.4	0.9	31	0.000

3.	I think my students do not seriously study the textbook materials.	2.9	1.7	25	0.000
4.	I skip certain topics in the textbook because they are less likely to be tested in the examination.	2.6	0.8	36	0.000
5.	My textbook (Traveller) is well-suited to developing the students' communicative competence.	2.7	1.5	23	0.000
6.	I rely on test papers and sample questions to prepare my students for the final test.	3.2	3.8	32	0.001
7.	I find interest in teaching textbook materials.	2.5	1.5	19	0.000
8.	If I teach the whole textbook (Traveller), my students will perform badly in the test.	2.4	0.9	31	0.000
9.	I use authentic materials along with the textbook for the students' practice of English language in the class.	2.5	1.5	19	0.000
10	I use modern aids and equipment to teach and practice English language skills.	2.9	1.7	25	0.000

1-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (B-1) was (19) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (B-1) According to the previous results we can say that our hypothesis is accepted

2-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (B-2) was (31) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (B-2) According to the previous results we can say that our hypothesis is accepted

3-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (B-3) was (25) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the

hypothesis (B-3) According to the previous results we can say that our hypothesis is accepted

4-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (B-4) was (36) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (B-4) According to the previous results we can say that our hypothesis is accepted

5-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (B-5) was (23) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (B-5) According to the previous results we can say that our hypothesis is accepted

6-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (B-6) was (32) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (B-6) According to the previous results we can say that our hypothesis is accepted

7-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (B-7) was (19) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (B-7) According to the previous results we can say that our hypothesis is accepted

8-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (B-8) was (31) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which

was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (B-8) According to the previous results we can say that our hypothesis is accepted

9-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (B-9) was (19) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (B-9) According to the previous results we can say that our hypothesis is accepted

10-The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (B-10) was (25) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (B-10) According to the previous results we can say that our hypothesis is accepted

Table No (4.55) Explains the study hypotheses of Teaching Methods

No.	Statement	mean	SD	Chi square	p-value
1.	I teach in a way that my students understand everything.	2.8	3.4	25	2.8
2.	I teach and explain the text in English.	2.5	1.5	12	2.5
3.	I encourage my students to ask questions during the class.	2.4	0.9	31	2.4
4.	I encourage my students to speak English in the class.	2.9	1.6	25	2.9
5.	I use Arabic along with English to make my students understand better.	2.6	0.7	36	2.6
6.	I teach whatever I think important to teach, no matter whether it is important or not for the test.	2.7	1.5	23	2.7
7.	My role as a teacher of English is to transmit knowledge to my students through explaining texts and giving examples.	2.8	3.4	25	2.8
8.	I do not make my students practice how to learn and speak English language but make them	2.5	1.5	12	2.5

	practice on how to answer questions in the test.				
9.	I believe that the test score in English is an appropriate indicator of a student's English ability.	2.4	0.9	31	2.4

The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (C-1) was (25) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (C-1). According to the previous results we can say that our hypothesis is accepted

Table No (4.56) Explains the study hypotheses of Classroom Tasks and Activities

No.	Statement	mean	SD	Chi square	p-value
1.	I ignore the task and activities that are not directly related to passing the test.	2.7	4.1	22	2.7
2.	I give model tests to the students to do better in the final test.	2.6	0.5	19	2.6
3.	I spend more time teaching grammar because I think grammar is more likely to be tested.	2.5	0.9	31	2.5
4.	I make my students practice and solve the questions of the past tests.	2.9	1.6	22	2.9
5.	The test discourages me to teach English language.	2.6	0.7	36	2.6
6.	I teach test-taking strategies, especially when the test date gets closer.	2.7	1.5	23	2.7

The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (D-1) was (22) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (D-1). According to the previous results we can say that our hypothesis is accepted

Table No (4.57) Explains the study hypotheses of EFL Skills and Elements

No.	Statement	mean	SD	Chi square	p-value
1.	The test influences my decision on	2.8	2.1	27	0.000

	which language skills are more important to be taught.				
2.	Listening is practiced in the class.	2.7	1.5	29	0.000
3.	Speaking is practiced in the class.	2.6	0.5	34	0.000
4.	Reading is practiced in the class.	2.4	1.6	27	0.000
5.	Writing is practiced in the class.	2.9	2.7	23	0.000

The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (E-1) was (27) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (E-1). According to the previous results we can say that our hypothesis is accepted

Table No (4.58) Explains the study hypotheses of Beliefs, Attitudes, and Perception Related to the Test and Teaching

No.	Statement	mean	SD	Chi square	p-value
1.	I feel pressure from my authority to improve my students' test score.	2.6	0.5	34	0.000
2.	My students can score good marks without improving their English language proficiency.	2.4	1.6	27	0.000
3.	I get feedback on my teaching from the test results.	2.9	2.7	23	0.000
4.	My students suffer from anxiety and tension for the test.	2.7	1.5	30	0.000
5.	The present testing system helps my students improve language proficiency.	2.8	0.6	24	0.000
6.	I could teach English better if there were no pressures for good results in the test.	2.7	1.5	29	0.000
7.	Tests influence my students' future career.	2.6	0.5	34	0.000
8.	I feel embarrassed if my students fail or perform badly in the test.	3.2	3.5	33	0.001

The calculated values of chi-square for the significance of the differences for the respondent's answers in the part No (F-1) was (34) which is greater than the tabulated value of chi-square at the degree of freedom (4) and the significant value level (5%) which

was (12.0). This indicates that, there are statistically significant differences at the level (5%) among the answers of the respondents, and also the calculated means are greater than the hypothesized mean (2.3) which support the respondents who agreed with the hypothesis (F-1) . According to the previous results we can say that our hypothesis is accepted

4.5 Summary

This chapter has covered the data analysis of the study which is about washback-based approach of the achievement test in assessing the learning process. This is done through a test (pre and posttest) to the first year secondary and a questionnaire to the teachers of EFL. Moreover, it showed the data tabulated in figures and tables. Then, interpretations were made from the collected data. Finally, the researcher has discussed the results of the study.

CHAPTER FIVE

MAIN FINDINGS, CONCLUSIONS, RECOMMENDATIONS AND SUGGESTIONS FOR FURTHER STUDIES

5.0. Introduction

This chapter presents the answers to research questions, including the discussion of main findings. Moreover, brief recommendation and suggestions for further studies were given at the end of the chapter. To fulfill the purpose of the study, the researcher applied two tools, namely, test (pre and posttests) and a questionnaire.

5.1 Main Findings

The results of this study proved that washback-based approach of the achievement test evaluate the learning process. The researcher has summarized following findings:

1- washback-based approach of the achievement test has positive impact on developing the EFL learning process.

2- Learners focus more on the test taking than on language learning strategies, this actually will lead to negative washback.

3- It is clear that tests influence different educational parts, particularly teachers and students in many ways.

4- Learning process can be improved by employing direct testing techniques.

5- Some teachers don't care about the syllabus and the curriculum while teaching their students.

6. The result of writing test shows that the performance of most study subjects in writing skill is weak. There for, writing skill as on one of the most important language competence is more challenging to them. Tables (4.1 and 4.6).

7- Teachers feel pressure to cover the syllabus before the final test.

8-Teachers are aware of the objectives of the syllabus and curriculum.

9- The present syllabus and curriculum can enhance EFL learning process.

10-The present teaching methods and techniques that adopted by teachers for teaching to test EFL student are not up to date.

- 11- Teachers almost teach every section in the text book (traveler) although some sections are unlikely to be tested.
- 12- Teachers tend to ignore subjects and activities that are not directly related to passing the test.
- 13- The achievement test is a major factor determining the future of EFL learners.
- 14- Test scores are used by policy makers in education as tools to control admission, promotion, placement and graduation.
- 15- Teachers feel embarrassed if their students fail or perform badly in the test.
- 16- Students could perform better if there were no pressures for good results in the test.
- 17- It is believed that the test score in English is an appropriate indicator of a student's English ability.
- 18- Washback is positive or negative depending on whether or not the test promoted the learners' language development.

5.2 Conclusion

This study deals with the analysis and evaluation of washback-based approach of the achievement test in assessing the learning process in five chapters.

In chapter one, the researcher has introduced a brief background of the study, statement of the research problems, significance of the study, objectives of the study, questions of the study, hypotheses of the study, method, population and sampling, instrument of data collection, limitations as well as tools.

Chapter two introduces the theoretical framework related to the washback- based approach of the achievement test and previous studies.

Chapter three presents the research methodology two instruments of data collection have been used (pretest, posttest for higher secondary learners and questionnaire for EFL teachers). (See appendix A and B).

In chapter four the results of the tests and the questionnaire have been analyzed statistically by using SPSS program. Then each of two questions and hypotheses have been discussed and answered respectively according to the results. Consequently, significant findings that show the positive influence of investigating the washback-based approach features on developing EFL learning process.

Finally, in chapter five, the research summary, main findings, recommendations and suggestions for further studies have been clarified. The study has added some insights to English language learners as well as their teachers particularly in the field of washback approach.

5.3 Recommendations

Considering the above-stated findings and results, the researcher recommends the following:

- 1-More attempts should be done to generate positive washback by means of examination reforms
- 2- EFL students in Saudi Arabia should be recommended to focus on language learning strategies and how to use it in real life contexts.
- 3- Teachers and learners should adopt more modern communicative approaches to language learning.
- 4- Students should be encouraged by teachers to study the language for communicative purposes not of only for passing the test.
- 5-Teachers should be responsible for helping their students overcome the obstacles they might face in language learning process.
- 6-Teachers are recommended to utilize from the test results to enhance the students' learning while at the same time not inhibiting students' motivation by cramming too much.

7- Learning more about washback in this environment has an important implication for teacher education programs that are designed to train new instructors, to introduce valuable classroom techniques and strategies.

8- It is hoped that these findings might be useful to teachers and learners of EFL in terms of having enough background about the term washback effects on teaching and learning process.

9-Students shouldn't get good marks without improving their English language proficiency.

10-The present testing system should help the students improve their language proficiency.

11.If the consequences of a particular test for teaching and learning are to be evaluated, the educational context in which the test takes place needs to be investigated. Whether the washback effect is positive or negative will largely depend on how it works and within which educational contexts.

12- Whether the influences of testing on teaching and learning are positive or negative is still debatable and needs to be studied further.

5.4 Suggestions for Further Studies

The present study presents the following suggestions for further researches:

1- Investigating more about washback-based approach of the achievement test is crucial.

2-Using modern aids and equipment to teach and practice English language skills is of great value.

3- Exploring up to date and effective teaching and learning methods to promote the positive washback and avoid the negative one.

3-Teaching plan should cover every section in the textbook although some sections are unlikely to be tested in the examination.

4-Testing the overall competence of the students in English such as listening, speaking, reading, and writing is required.

- 5- Special training courses are needed in the concept of washback to develop good exams.
- 6- The results of this study seem to indicate that there is a lack of research regarding washback approach and how this gap could be closed. Therefore, it deserves more attention.
7. Washback should be one of the most important criterion for developing and evaluating language test and teaching process.
8. The policy makers need to take into consideration the power of beliefs in their planning and strategies.
9. Understanding what types of beliefs lead to desired outcomes and what types of beliefs are stumbling blocks would be of use to help promote positive washback and minimize undesirable negative washback.

BIBLIOGRAPHY

- Alderson, J.C. and J. Banerjee. 2001. "Language testing and assessment (Part 1)". *Language Teaching* 34, 4: 213-236 *ELT Journal*.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Alderson, J. C. (1987). *Innovation in language testing. Can the micro-computer help? Special report number 1, Language Testing Update*. Lancaster: University of Lancaster.
- Alderson, J. C. (1986). *Innovations in Language Testing? Language Testing: Proceedings of the IUS/NFER Conference*. Philadelphia: NFER-Nelson
- Andrews, S., Fullilove, (2004). *Targeting washback – a case study*. Oxford: Oxford University Press
- Andrews, S., Fullilove, J., & Wong, Y. (2002). *Targeting washback: A case study*. *System*, 30, 207-223. *ELT Journal*.
- Andrews, S., & Fullilove, J. (1994). *Assessing spoken English in public examinations-Hong Kong*: Chinese University Press.
- Bachman, L.F. & Palmer, A.S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Bailey, K. M. (1996). *Working for washback: A review of the washback concept in language testing. Language Testing*. Cambridge: Cambridge University Press.
- Biggs, J. B. (1994). *Assumptions underlying new approaches to educational assessment*. *Curriculum Forum*, 4, 1-22. Unpublished MA thesis, University of Lancaster.
- Brown, H. D. (2004). *Language Assessment, Principles and Classroom Practices*. USA: Longman
- Buck, G. (1988). *Testing listening comprehension in Japanese university entrance examinations*. *JALT Journal*, 10, 12-42.

- Burrows, C. (2004). *Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program*. Mahwah, NJ: Lawrence Erlbaum Assoc.
- Burrows, C. (1998) *Searching for washback: an investigation of the impact on teachers of the implementation into the Adult Migrant English Program of the assessment of the Certificates in Spoken and Written English. Ph.D. thesis*. Sydney: Macquarie University, Sydney
- Chapman, D. W., & Synder, C. W. (2000). *Can high-stakes national testing improve instruction: Reexamining conventional wisdom*. *International Journal of Educational Development*, 20, 457-474.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. *Studies in language testing*, 21. Cambridge: Cambridge University Press.
- Cheng, L. (2004). *The washback effect of a public examination change on teacher's perceptions toward their classroom teaching*. New York: Cambridge University Press.
- Cheng, L. (2002). *The washback effect on classroom teaching of changes in public examinations*. New Haven, CT: Yale University Press.
- Cheng, L. (2000). *Distance language teacher education: New challenges for Hong Kong*. *Open Learning*, 15(1), 5-16.
- Cheng, L., & Falvey, P. (2000) *What works? The washback effect of a new public examination on teachers' perspectives and behaviors in classroom teaching*. *Curriculum Forum*, 9(2), 10-33
- Cheng, L. (1998). *Impact of a public English examination change on students' perceptions and attitudes toward their English learning*. *Studies in Educational Evaluation*, 24(3), 279-301.
- Cheng, Sh. (1997). *On College English Test Band IV and Band VI*. *Journal of Guangzhou Normal University*
- Cohen, L., & Manion, L. (1994). *The impact of EFL testing on EFL education in Korea*. *Language Testing*, 25(1), 39-62.
- Davies, P. M. & Pearse, E. (2001) *Success in English Teaching*. Oxford: Oxford University Press.

- Davies, A., A. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.
- Davies, A. (1985). *Follow my leader: Is that what language tests do?* Oxford: Pergamon Press.
- Davison, C. (2008). *Using summative assessments for formative purposes: The ultimate justification for learners and teachers*. Hangzhou, China University Press
- Eckstein, (1993). *Secondary school examinations: international perspectives on policies and practice*, New Haven, CT: Yale University Press
- Fredericksen, N. (1984). *The real test bias: Influences on testing and teaching*. *American Psychologist*, 39, 193-202.
- Freeman, D. (1996). *Voices from the language classroom: Qualitative research in second language education*. Cambridge: Cambridge University Press.
- Gates, S. (1995). *Exploiting washback from standardized tests*. Tokyo: Japan Association for language Teaching. Language in India www.languageinindia.com
- Genesee and John A. Upshur. (2001) *Language Evaluation and Assessment* Cambridge: Cambridge University Press.
- Gleason, H.A. (1961). *An Introduction to descriptive Linguistics*. New York: Holt, Rinehart and Winston
- Green, A. (2007). *IELTS washback in context*. Cambridge: Cambridge University Press.
- Green, A. (2006). *Watching for Washback: Observing the Influence of the International English Language Testing System Academic Writing Test in the Classroom*. *Language Assessment Quarterly*, 3(4), 333-368.
- Green, D. Z. (1985). *Developing measures of communicative proficiency: A test for French immersion students in grades 9 and 10*. Ottawa: University of Ottawa Press

Hamp-Lyons, L. (1996). *Ethical test preparation: The case of TOEFL. Paper presented at the Language Testing Research Colloquium, Chicago.* Chicago University Press.

Hawkey, R. (2006). *Teacher and learner perceptions of language learning activity.* English Language Teaching Journal, 60(3), pp. 242-252.

Hayes, B. and J. Read, (2004). *Washback in language testing: Research contexts and method.* Mahwah: Lawrence Erlbaum.

Heaton, J.B. (1988). *Writing English Language Tests.* London: Longman

Henning, G. (1987). *A guide to language testing." Development, evaluation, research,* New York: Newbury House

Heyneman, S. P. (1987). *Uses of examinations in developing countries: Selection, research, and education sector management.* International Journal of Educational Development 7(4), 251-263.

Hsu, Hui-Fen (2009). *The impact of implementing English proficiency tests as a graduation requirement at Taiwanese universities of technology. Ph.D. thesis.* University of York. Retrieved January 27, 2010 from -etheses.whiterose.ac.uk

Hughes, A. (2003). *Testing for Language Teachers, 2nd Edition.* Cambridge: Cambridge University Press.

Hughes, A. (1989). *Testing for language teachers.* Cambridge: Cambridge University Press.

Hughes, A. (1988). *Introducing a needs-based test of English language proficiency into an English medium university in Turkey.* Modern English Publications.

Johnson, K. (1992). *The instructional decisions of pre-service ESL teachers: New directions for teacher preparation programs.* Hong Kong: City Polytechnic of Hong Kong.

Lam, H. P. (1994). *Methodology washback—an insider's view.* Hong Kong: University of Hong Kong.

Madaus, G. (1988). *Testing and the curriculum.* Chicago: University of Chicago Press, p. 83-121.

- Mehrens, W.A.(1991). “*Defensible/indefensible instructional preparation for high stakes achievement tests: an exploratory dialogue*”. ERIC Document Reproduction Service, No. ED 334202.
- Messick, S. (1996). *Validity and washback in language testing*. Language Testing, 13(3), 241-256.
- Morris, P. (1990). *Curriculum development in Hong Kong*. Hong Kong: Faculty of Education. The University of Hong Kong.
- Morrow, K. (1991). *Evaluating communicative tests*. In S. Anivan (Ed.), *Current developments in language testing* (pp. Singapore: Regional Language Centre
- Oller, J. W., Jr. (1979). *Language tests at school: A pragmatic approach*. London: Longman.
- Pearson, I. (1988). *Tests as levers of change*. London: Modern English Publications in association with the British Council.
- Pierce, B. N. (1992). *Demystifying the TOEFL Reading Test*. *TESOL Quarterly*, 26(4), 665-689.*The merits of measurement-driven instruction*. Phi Delta Kappa, 68, 679– 682
- Qi, L. (2004). *Washback in language testing: Research contexts and methods* (pp. 171-190). Mahwah, NJ: Lawrence Erlbaum Associates.
- Richards, J. (2001) *Curriculum Development in Language Teaching*, Cambridge: Cambridge University Press
- Sato K and Kleinsasser RC (1999) *Communicative Language Teaching (CLT): Practical Understandings*. India. The Modern Language Journal, 83(4),494-517.
- Shepard, L.A. (1991). *Psychometricians’ beliefs about learning*. *Educational Researcher*. ELT Journal, 20(6), 2–16.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). *Test impact revisited: Washback effect over time*. Language Testing, 13(3), 298-317.
- Shohamy, E. (1993). *A collaborative/diagnostic feedback model for testing foreign languages*. Alexandria, VA: TESOL Publications.
- Shohamy, E. (1992). *Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning*. *Modern Language*

Journal, 76(4), 513521. Dubuque, IA: Kendall Hunt Publishing Company.

Shohamy,(1986). *Introducing a new comprehensive test of oral proficiency*. *ELT Journal*, 40(3), 212-220.

Shulman, L.S. (1986a). *Paradigms and research programmes in the study of teaching: a contemporary perspective*, in M.C. Wittrock (Ed.) New York: Macmillan.

Smith, M. L. (1991). *Put to the test: The effects of external testing on teachers*. *Educational Researcher*. Arizona, USA: Arizona State University Press.

Spolsky, B. (1995). *What does it mean to know how to use a language? An essay on the theoretical basis of language testing*. India www.languageinindia.com 16:9 September 2016 Md. Enamul Hoque Ph.D

Spratt, M. (2005). *Washback and the classroom: The implications for teaching and learning of studies of washback from exams*. Washington, DC: TESOL Publications.

Swain, M. (1985). *Large-scale communicative testing: A case study*. In Y. P. Lee, A. C. Y. Fok, R. Lord, & G. Low (Eds.), *Oxford*: Pergamon Press.

Tan, M. (2008). *Bilingual high-stakes mathematics and science exams in Malaysia: pedagogical and linguistic issues*. Paper presented at the *Language Testing and Research Colloquium 2008*, Hangzhou.

-Tzagari, D. (2007). *Investigating the Washback effect of the FCE examination (Cambridge/ESOL) in Greece*. Greece, UCLES.

Turner, C.E. (2008). *The specificity of the “research approach” in classroom studies: Probing the predictability of washback through teacher conceptual and instrumental evidence in Quebec high schools*. Paper presented at the *Language Testing and Research Colloquium 2008*, Hangzhou.

Valette, R.M. (1994). *Teaching, testing, and assessment: conceptualizing the relationship*. In C. Hancock (Ed.), *Teaching, Testing, and Assessment Making the Connection*. (pp. 1-42). Lincolnwood, Illinois. National Textbook Company.

- Vernon, P.E. (1956). *The measurement of abilities (2nd ed.)*. London: University of London Press.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching*. Cambridge, UK: Cambridge University Press.
- Wall, D. (1999). *The impact of high-stakes examinations on classroom teaching: A case using insights from testing and innovation theory*. Unpublished doctoral dissertation. England: The University of Lancaster.
- Wall, D. (1998). *Impact and washback in language testing*. Encyclopedia of language and education, 7, 291-302. Dordrecht: Kluwer Academic Publishers
- Wall, D. (2000). *The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled?* System, 28, 499-509.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching*. Cambridge, UK: Cambridge University Press.
- Wall, D. (1997). *Impact and washback in language testing*. In C. Clapham & D. Corson (Eds.), *Encyclopedia of Language and Education*. (pp. 291-302). The Netherlands: Kluwer Academic Publishers.
- Wall, D. & Alderson, C. (1996). *Validation in language testing (pp. 194-221)*. Philadelphia: Multilingual Matters.
- Wall, D., & Alderson, J. C. (1993). *Examining washback: The Sri Lankan impact study*. *Language Testing*, 10(1), 41-69. Language in India www.languageinindia.com
- Watanabe, Y. (2004a). *Methodology in washback studies*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Watanabe, Y. (1996). *Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research*. *Language Testing*, 13(3), 318-333.
- Wesche, M. B. (1983). *Communicative language testing in a second language*. *Modern Language Journal*, 67(1), 41-55. Language in India www.languageinindia.com

Wesche, M. B. (1987). *Second language performance testing: The Ontario Test of ESL as an example*. *Language Testing*, 4(1), 28-47.

Wesdorp, H. (1982). *Backwash effects of language testing in primary and secondary education*. Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam. Amsterdam.

Woods, D. (1996). *Teacher Cognition in Language Teaching*. Cambridge: Cambridge University Press.

Appendix(A)
Students' Test

Kingdom of Saudi Arabia

Ministry of Education

The General Directorate of Education in the Eastern Province

Qatif Ahliya Schools

Stage: Secondary

Grade:1

Academic Year: 2017/2018-AD -1438/1439-H

Term:1

Date:

Time: 3 hours

Student Name:

Total Mark: _____

50

Q1. Reading Comprehension:

(10 marks)

Read the Following Text and Then Answer the Questions Below:

The Internet is part of our everyday life. It helps us a lot: we find information, we send e-mails, we buy things, etc. and it's easy to use. But be careful! There are lots of dangers, like spamming and phishing.

Spamming is sending unwanted e-mails like advertising material over the Internet to a large number of people. You can get lots of spam even on your fax machine or on your mobile phone as SMS! Junk e-mail is a type of spam. These e-mails try to make you buy things.

Phishing is a very big problem, because the 'phishers' want to get your personal information, like your bank or credit card numbers so they can use them. Phishers make a fake website that looks exactly like another website and send the link to people. People see it and think that it's from a bank, for example, and fill in their personal information. Phishing is common on social networking websites. So, be careful!

(A). Answer the Following Questions:

1. How does the Internet help us?

.....

2. What are the dangers of the Internet?

.....

3. What are Junk e-mails?

.....

4. Why is Phishing a very big problem?

.....

5. What is Spamming?

.....

P.T.O.

(B). Answer with True (T) or False (F):

- 1. The Internet isn't part of our everyday life. ()
- 2. People use the Internet for different reasons. ()
- 3. You can get lots of spam even on your school books. ()
- 4. Phishers want to get your personal information. ()
- 5. Phishers make fake websites. ()

Q2. Grammar:

(10 marks)

(A) Choose the correct answer:

- 1. This clock isn't..... Is it yours?
a. ours b. our c. my d. he
- 2. **A:** mobile is this? **B:** It's Jerry's.
a. Who's b. What c. Whose d. Why
- 3. I don't workThursday.
a. on b. at c. in d. for
- 4. It is the ball.
a. boy b. boy's c. boys d. boy'
- 5. ITV every day .
a. watches b. am watching c. have watched d. watch

(B) Rearrange the following sentences:

- 1. They / always / TV? / study / in front of
.....
- 2. walking / hates / to work / Fatima
.....
- 3. enjoy / talking / My sisters / on the phone
.....
- 4. gets up / early / Mohammed / in the morning
.....
- 5. e. mails / a lot of / I / receive / every day.
.....

P.T.O.

Q3. Vocabulary:

(10 marks)

(A) Underline the correct words:

- 1.They're really (**hungry** / **busy**). They can't wait till lunch break.
- 2.Adam can't work as a secretary because he can't (**use** / **type**) a computer.
- 3.Hassan's flat is very (**depressed** / **convenient**) because it's near the mall.
4. We love Paris, (**special** / **especially**) in the spring.
5. I want to have a (**talk** / **word**) with you tonight.

(B) Match the following:

1. do () a. shopping at the mall
2. go () b. dinner
3. take up () c. homework
4. hang out () d. with friend
5. have () f. sport

Q4. Conversation:

(10 marks)

Choose from the list to complete the dialogue:

(watching / live / meet / go / playing / sisters / got / doesn't / am / same)

Alex: Hello. I'm Alex.

Tom: Hi! I'm Tom. Are you new?

Alex: Yes, IWe're in theclass.

Tom: Oh! Nice toyou. Where's your house?

Alex: Inear the school. What about you?

Tom: Me, too. What are your hobbies?

Alex: Well, I play football and I swimming at the weekend.

Do you like sport?

Tom: Not really. I likecomputer games andTV.

Alex: Have you got any brothers or?

Tom: Yes, I'veone brother. But he's three years old so he
..... go to school.

Q5.Writing:

(10marks)

Write a Short Paragraph About You:

- Write your name / age / occupation / nationality.
- Write about your favorite food/ drink/ school subject / sport.
- Write about things you like doing and things that you don't like.
- Write about what you are like / what you look like / the clothes you like wearing
- Write about your daily routine / free- time activities / hobbies.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

End of questions

Best wishes

APENDIX (B)
Teachers' Questionnaire

Dear EFL Teacher,

This Questionnaire has been developed for the purpose of a research project in the Department of English at Sudan University of Science and Technology, Khartoum, Sudan. The statements here are related to the “Washback based Approach of the Achievement Test in Assessing the EFL Learning Process”. The researcher gives you full assurance that the information will be used only for the research purpose and will be strictly kept confidential.

Thank you for cooperation!

Name: Ibrahim Ahmed Eisa Abdallah
PhD. candidate (Applied Linguistics- English Language)
e-mail : Abdallahibrahim659@yahoo.com

Guidelines

The survey questionnaire in hand consists of two parts:

Part One: Background / Personal information of respondents.

Directions: Please answer thoroughly and tick what suits you.

1. Name (optional) :
2. Job title: a. Teacher () b. Supervisor () c. Group leader ()
3. Place of Work:
4. Teaching experience: () years .
5. Age: () years .
6. Sex: Male () Female ()
7. Education and qualifications:(Please tick the highest degree obtained) .
 - a. Diploma () b. Post-Secondary Diploma / TEFL ()
 - c. BA in English ()
 - d. BA in English & Education () e. MA in English & Education ()
 - f. Others (please specify)

Part Two: Questionnaire.

Directions: Put a tick mark (✓) in the box next to each item, which best expresses your opinion:

A. Curriculum and Syllabus

No.	Statement	Agree	Neutral	Disagree
1.	I am aware of the objectives of the syllabus and curriculum.			
2.	The present syllabus and curriculum can enhance EFL teaching and learning.			
3.	I teach every section in the textbook (<i>Traveller</i>) although some sections are unlikely to be tested in the examination.			
4.	I do not care about the syllabus and curriculum while teaching my students.			
5.	I feel pressure to cover the syllabus before the final test.			
6.	The test tests the overall competence of my students in English (listening, speaking, reading, and writing).			
7.	I give more attention to teaching to the syllabus opposed to practicing the test items.).			

B. Textbooks and Materials

No.	Statement	Agree	Neutral	Disagree
1.	I follow and communicate the lesson objectives with the students while planning my lessons.			
2.	The textbook (<i>Traveller</i>) covers exercises that the curriculum has claimed.			
3.	I think my students do not seriously study the textbook materials.			
4.	I skip certain topics in the textbook because they are less likely to be tested in the examination.			
5.	My textbook (<i>Traveller</i>) is well-suited to developing the students' communicative competence.			
6.	I rely on test papers and sample questions to prepare my students for the final test.			
7.	I find interest in teaching textbook materials.			
8.	If I teach the whole textbook (<i>Traveller</i>), my students will perform badly in the test.			
9.	I use authentic materials along with the textbook for the students' practice of English language in the class.			
10.	I use modern aids and equipment to teach and practice English language skills.			

C. Teaching Methods

No.	Statement	Agree	Neutral	Disagree
1.	I teach in a way that my students understand everything.			
2.	I teach and explain the text in English.			
3.	I encourage my students to ask questions during the class.			
4.	I encourage my students to speak English in the class.			
5.	I use Arabic along with English to make my students understand better.			
6.	I teach whatever I think important to teach, no matter whether it is important or not for the test.			
7.	My role as an English teacher is to transmit knowledge to my students through explaining texts and giving examples.			
8.	I do not make my students practice how to learn and speak English			

	language but make them practice on how to answer questions in the test.			
9.	I believe that the test score in English is an appropriate indicator of a student's English ability.			

D. Classroom Tasks and Activities

No.	Statement	Agree	Neutral	Disagree
1.	I ignore the task and activities that are not directly related to passing the test.			
2.	I give model tests to the students to do better in the final test.			
3.	I spend more time teaching grammar because I think grammar is more likely to be tested.			
4.	I make my students practise and solve the questions of the past tests.			
5.	The test discourages me to teach English language.			
6.	I teach test-taking strategies, especially when the test date gets closer.			

E. EFL Skills and Elements

No.	Statement	Agree	Neutral	Disagree
1.	The test influences my decision on which language skills are more important to be taught.			
2.	Listening is practiced in the class.			
3.	Speaking is practiced in the class.			
4.	Reading is practiced in the class.			
5.	Writing is practiced in the class.			

F. Beliefs, Attitudes, and Perception Related to the Test and Teaching

No.	Statement	Agree	Neutral	Disagree
1.	I feel pressure from my authority to improve my students' test score.			
2.	My students can score good marks without improving their English language proficiency.			
3.	I get feedback on my teaching from the test results.			
4.	My students suffer from anxiety and tension for the test.			
5.	The present testing system helps my students improve language proficiency.			
6.	I could teach English better if there were no pressures for good results in the test.			
7.	Tests influence my students' future career.			
8.	I feel embarrassed if my students fail or perform badly in the test.			

Free Space: Feel free to use the following space for any further suggestions and comments:

.....

APENDIX (C)

The Validation Jury

1. Dr. Ahmed Khidr Ahmed Othman.

Tel: 00966553549272

Tabouk University. Faculty of Education and Arts. English Language
Department

2. Dr. Nassir Ali Othman Saty

Tel: 00249908124644

Tabouk University. Faculty of Education and Arts. English Language
Department